



(86) Date de dépôt PCT/PCT Filing Date: 2005/11/04
(87) Date publication PCT/PCT Publication Date: 2006/05/18
(85) Entrée phase nationale/National Entry: 2007/04/25
(86) N° demande PCT/PCT Application No.: US 2005/039970
(87) N° publication PCT/PCT Publication No.: 2006/052731
(30) Priorité/Priority: 2004/11/05 (US60/625,442)

(51) Cl.Int./Int.Cl. *C12Q 1/68* (2006.01),
C12N 15/10 (2006.01), *G01N 33/574* (2006.01)

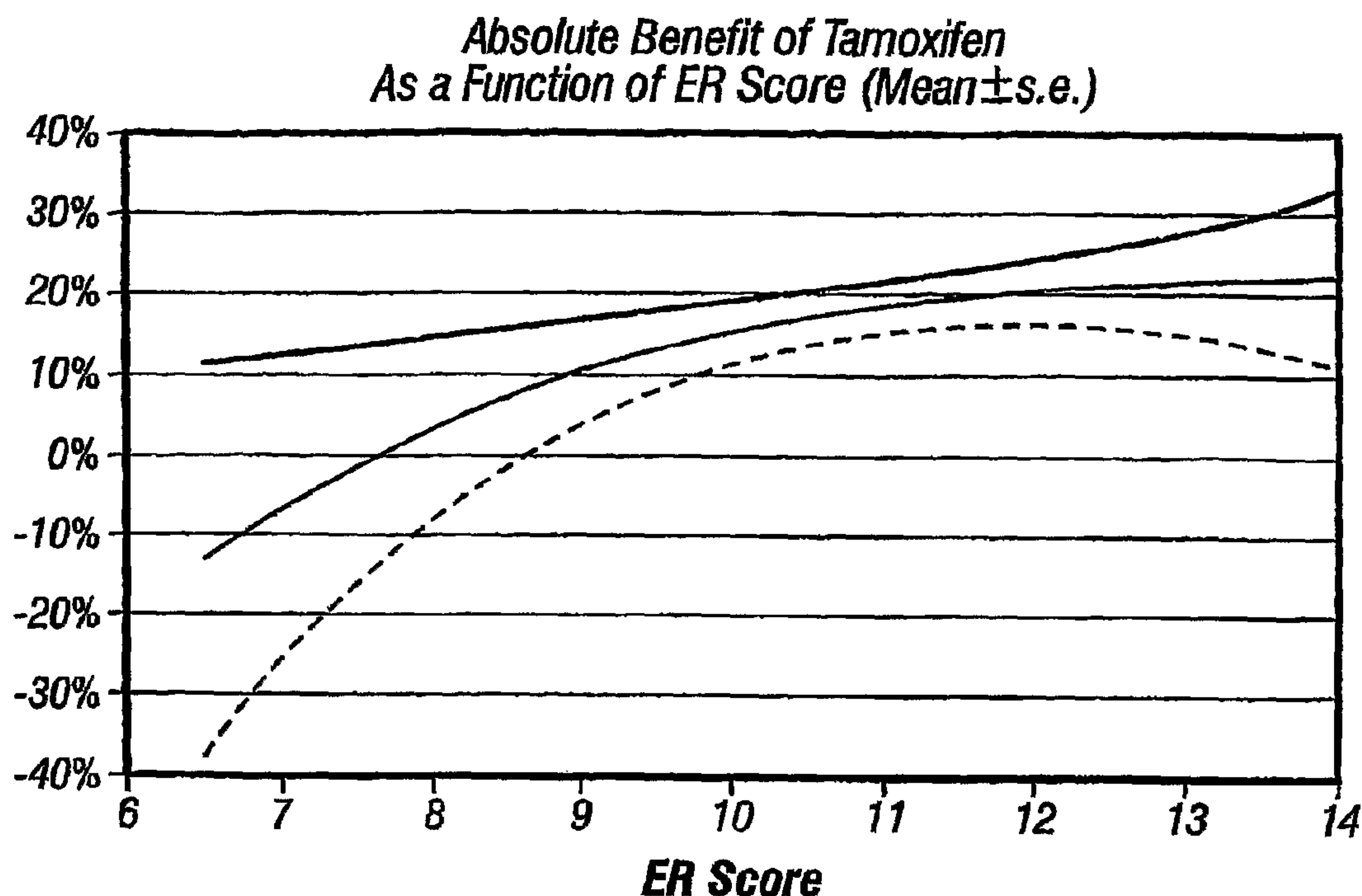
(71) Demandeurs/Applicants:
GENOMIC HEALTH, INC., US;
NSABP FOUNDATION, INC., US

(72) Inventeurs/Inventors:
BAKER, JOFFRE B., US;
BRYANT, JOHN L., US;
PAIK, SOONMYUNG, US;
SHAK, STEVEN, US

(74) Agent: SMART & BIGGAR

(54) Titre : INDICATEURS MOLECULAIRES DE PRONOSTIC DE CANCER DU SEIN ET PREDICTION DE REPONSE
DE TRAITEMENT

(54) Title: MOLECULAR INDICATORS OF BREAST CANCER PROGNOSIS AND PREDICTION OF TREATMENT
RESPONSE



(57) Abrégé/Abstract:

The present invention relates to quantitative molecular indicators that can guide clinical decisions in breast cancer, such as estrogen receptor (ESR1)-positive, lymph node-negative breast cancer. In particular, the invention concerns certain genes, the varied expression of which indicates the likelihood of recurrence of surgically resected breast cancer in patients who are not treated with a therapeutic agent in the adjuvant setting. In addition, the invention concerns the use of quantitative measurement of the expression of certain genes, including the ESR1 gene, that measures as a continuous variable, to determine (a) the likelihood of a beneficial response to the anti-estrogen therapeutic agent, such as tamoxifen; and (b) the potential magnitude of beneficial response to chemotherapy.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
18 May 2006 (18.05.2006)

PCT

(10) International Publication Number
WO 2006/052731 A2(51) International Patent Classification: **Not classified**

(21) International Application Number:

PCT/US2005/039970

(22) International Filing Date:

4 November 2005 (04.11.2005)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/625,442 5 November 2004 (05.11.2004) US

(71) Applicants (for all designated States except US): **GENOMIC HEALTH, INC.** [US/US]; 301 Penobscot Drive, Redwood City, CA 94063 (US). **NSABP FOUNDATION, INC.** [US/US]; Four Allegheny Center, 5th Floor, Pittsburgh, PA 15212 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BAKER, Joffre, B.** [US/US]; 1400 Avery Street, Montara, CA 94037 (US). **BRYANT, John, L.** [US/US]; 9705 Cole Road, Allison Park, PA 15101 (US). **PAIK, Soonmyung** [US/US]; 100 Anderson Street, Pittsburgh, PA 15201 (US). **SHAK, Steven** [US/US]; 648 Fairway Circle, Hillsborough, CA 94010 (US).(74) Agent: **FOSTER, David**; Heller Ehrman LLP, 275 Middlefield Road, Menlo Park, CA 94025 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

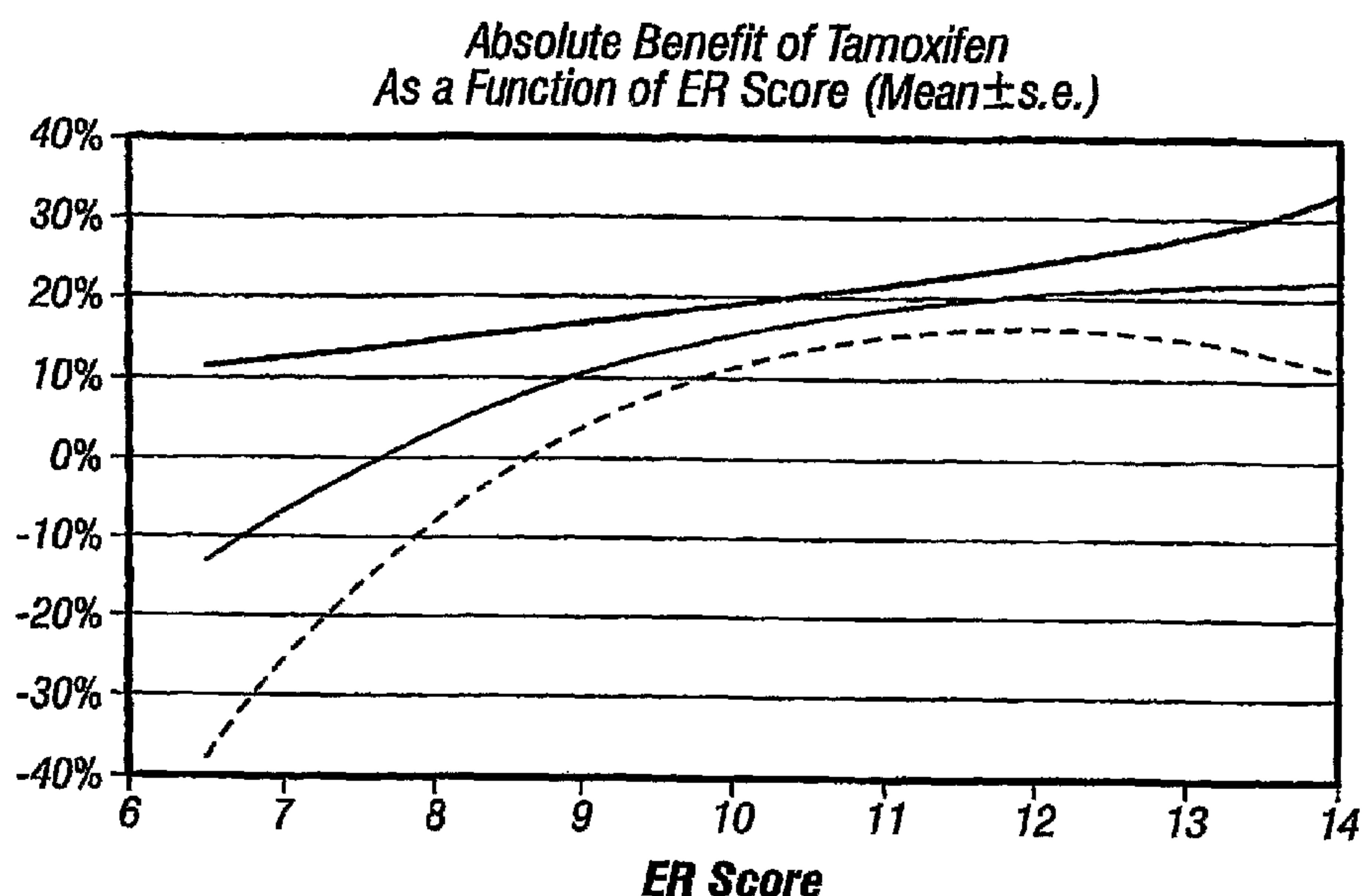
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

[Continued on next page]

(54) Title: MOLECULAR INDICATORS OF BREAST CANCER PROGNOSIS AND PREDICTION OF TREATMENT RESPONSE



(57) Abstract: The present invention relates to quantitative molecular indicators that can guide clinical decisions in breast cancer, such as estrogen receptor (ESR1)-positive, lymph node-negative breast cancer. In particular, the invention concerns certain genes, the varied expression of which indicates the likelihood of recurrence of surgically resected breast cancer in patients who are not treated with a therapeutic agent in the adjuvant setting. In addition, the invention concerns the use of quantitative measurement of the expression of certain genes, including the ESR1 gene, that measures as a continuous variable, to determine (a) the likelihood of a beneficial response to the anti-estrogen therapeutic agent, such as tamoxifen; and (b) the potential magnitude of beneficial response to chemotherapy.

WO 2006/052731 A2

WO 2006/052731 A2



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

MOLECULAR INDICATORS OF BREAST CANCER PROGNOSIS AND PREDICTION OF TREATMENT RESPONSE

Field of the Invention

The present invention relates to quantitative molecular indicators that can guide
5 clinical decisions in breast cancer, such as estrogen receptor (ESR1)-positive, lymph node-
negative breast cancer. In particular, the invention concerns certain genes, the varied
expression of which indicates the likelihood of recurrence of surgically resected breast cancer
in patients who are not treated with a therapeutic agent in the adjuvant setting. In addition,
the invention concerns the use of quantitative measurement of the expression of certain
10 genes, including the ESR1 gene, that measure as a continuous variable, to determine (a) the
likelihood of a beneficial response to the anti-estrogen therapeutic agent, such as tamoxifen;
and (b) the potential magnitude of beneficial response to chemotherapy.

Description of the Related Art

15

Gene expression studies

Oncologists have a number of treatment options available to them, including different
combinations of chemotherapeutic drugs that are characterized as “standard of care,” and a
number of drugs that do not carry a label claim for the treatment of a particular cancer, but
20 for which there is evidence of efficacy in that cancer. Best likelihood of good treatment
outcome requires that patients at highest risk of metastatic disease be identified and assigned
to optimal available cancer treatment. In particular, it is important to determine the
likelihood of patient response to “standard of care” therapeutic drugs, such as
cyclophosphamide, methotrexate, 5-fluorouracil, anthracyclines, taxanes, and anti-estrogen
25 drugs, such as tamoxifen, because these have limited efficacy and a spectrum of often severe
side effects. The identification of patients who are most or least likely to need and respond to
available drugs thus could increase the net benefit these drugs have to offer, and decrease net
morbidity and toxicity, via more intelligent patient selection.

Currently, diagnostic tests used in clinical practice are single analyte, and therefore do
30 not capture the potential value of knowing relationships between dozens of different markers.
Moreover, diagnostic tests are often based on immunohistochemistry, which is not
quantitative. Immunohistochemistry often yields different results in different laboratories, in

-2-

part because the reagents are not standardized, and in part because the interpretations are subjective. RNA-based tests, while potentially highly quantitative, have not been used because of the perception that RNA is destroyed in tumor specimens as routinely prepared, namely fixed in formalin and embedded in paraffin (FPE), and because it is inconvenient to
5 obtain and store fresh tissue samples from patients for analysis.

Over the last two decades molecular biology and biochemistry have revealed hundreds of genes whose activities influence the behavior of tumor cells, their state of differentiation, and their sensitivity or resistance to certain therapeutic drugs. However, with a few exceptions, the status of these genes has not been exploited for the purpose of routinely
10 making clinical decisions about drug treatments. In the last few years, several groups have published studies concerning the classification of various cancer types by microarray gene expression analysis of thousands of genes (see, e.g. Golub et al., *Science* 286:531-537 (1999); Bhattacharjee et al., *Proc. Natl. Acad. Sci. USA* 98:13790-13795 (2001); Chen-Hsiang et al., *Bioinformatics* 17 (Suppl. 1):S316-S322 (2001); Ramaswamy et al., *Proc. Natl. Acad. Sci.*
15 *USA* 98:15149-15154 (2001); Martin et al., *Cancer Res.* 60:2232-2238 (2000); West et al., *Proc. Natl. Acad. Sci. USA* 98:11462-114 (2001); Sorlie et al., *Proc. Natl. Acad. Sci. USA* 98:10869-10874 (2001); Yan et al., *Cancer Res.* 61:8375-8380 (2001)). However, these studies have not yet yielded tests routinely used in clinical practice, in large part because microarrays require fresh or frozen tissue RNA and such specimens are not present in
20 sufficient quantity to permit clinical validation of identified molecular signatures.

In the past three years, it has become possible to profile gene expression of hundreds of genes in formalin-fixed paraffin-embedded (FPE) tissue using RT-PCR technology. Methods have been described that are highly sensitive, precise, and reproducible (Cronin et al., *Am. J. Pathol.* 164:35-42 (2004)). Because thousands of archived FPE clinical tissue
25 specimens exist with associated clinical records, such as survival, drug treatment history, etc., the ability to now quantitatively assay gene expression in this type of tissue enables rapid clinical studies relating expression of certain genes to patient prognosis and likelihood of response to treatments. Using data generated by past clinical studies allows for rapid results because the clinical events are historical. In contrast, for example, if one wished to carry out
30 a survival study on newly recruited cancer patients one would generally need to wait for many years for statistically sufficient numbers of deaths to have occurred.

Breast cancer prognosis and prediction

Breast cancer is the most common type of cancer among women in the United States, and is the leading cause of cancer deaths among women ages 40 - 59.

Currently only a few molecular tests are routinely used clinically in breast cancer. Immunohistochemical assays for estrogen receptor (ESR1) and progesterone receptor (PGR) proteins are used as a basis for selection of patients to treatment with anti-estrogen drugs, such as tamoxifen (TAM). In addition, ErbB2 (Her2) immunochemistry or fluorescence in situ hybridization (which measure protein and DNA, respectively) are used to select patients with the Her2 antagonist drugs, such as trastuzumab (Herceptin®; Genentech, Inc., South San Francisco, CA).

Because current tests for prognosis and for prediction of response to chemotherapy are inadequate, breast cancer treatment strategies vary between oncologists (Schott and Hayes, *J. Clin. Oncol.* PMID 15505274 (2004); Hayes, *Breast* 12:543-9 (2003)). Generally, lymph node negative patients whose tumors are found to be ESR1 positive are treated with an anti-estrogen drug, such as TAM, and patients whose tumors are found to be ESR1 negative are treated with chemotherapy. Often, ESR1 positive patients are also prescribed chemotherapy in addition to anti-estrogen therapy, accepting the toxic side effects of chemotherapy in order to modestly decrease the risk of cancer recurrence. Toxicities include, neuropathy, nausea and other gastrointestinal symptoms, hair loss and cognitive impairment. Recurrence is to be feared because recurrent breast cancer is usually metastatic and poorly responsive to treatment. Clearly, a need exists to identify those patients who are at substantial risk of recurrence (i.e., to provide *prognostic* information) and likely to respond to chemotherapy (i.e., to provide *predictive* information). Likewise, a need exists to identify those patients who do not have a significant risk of recurrence, or who are unlikely to respond to chemotherapy, as these patients should be spared needless exposure to these toxic drugs.

Prognostic factors differ from treatment predictive factors in breast cancer. Prognostic factors are those variables related to the natural history of breast cancer, which influence the recurrence rates and outcome of patients once they have developed breast cancer. Clinical parameters that have been associated with a worse prognosis include, for example, lymph node involvement, increasing tumor size, and high grade tumors. Prognostic factors are frequently used to categorize patients into subgroups with different baseline relapse risks. In contrast, treatment predictive factors are variables related to the likelihood

of an individual patient's beneficial response to a treatment, such as anti-estrogen or chemotherapy, independent of prognosis.

There is a great need for accurate, quantitative prognostic and predictive factors that can assist the practicing physician to make intelligent treatment choices, adapted to a particular patient's needs, based on well founded risk-benefit analysis.

Summary of the Invention

In one aspect, the invention concerns a method for the prognosis of disease outcome in a breast cancer patient, comprising

10 (a) quantitatively determining, in a biological sample comprising cancer cells obtained from the patient, the value of one or more of the following variables:

(i) Proliferation Group Score;

(ii) Invasion Group Score;

(iii) Proliferation Group Threshold Score; and

15 (iv) the expression levels of the RNA transcripts of one or more of the following individual genes CCNB1, BIRC5, MYBL2, PGR, STK6, MKI67, GSTM1, GAPD, RPLPO, and MMP11, or their expression products; wherein

(b1) for every unit of an increase in the value of one or more of (i) - (iii) and/or of the RNA transcript(s) of one or more of the individual genes CCNB1, BIRC5, MYBL2, 20 STK6, MKI67, GAPD, and MMP11, or the corresponding expression product(s), the patient is identified to have a proportionately increased risk of poor disease outcome; and

(b2) for every unit of increased expression level(s) of the RNA transcript(s) of one or more of the individual genes PGR, GSTM1, and RPLPO, or the corresponding expression product(s), the patient is identified to have a proportionately decreased risk of poor disease 25 outcome,

wherein

Proliferation Group Score = (BIRC5 + MKI67 + MYBL2 + CCNB1 + STK6)/5;

Invasion Group Score = (CTSL2 + MMP11)/2;

Proliferation Group Threshold Score equals 6.5 if the Proliferation Group Score is less than 6.5; and is identical with the Proliferation Group Score, if the Proliferation Group Score is 6.5 or more, wherein

the gene symbols in the equations represent the expression levels of the RNA transcripts of the respective genes, or their expression products; and

wherein every individual gene or gene present in any of the variables can be substituted by another gene that coexpresses with said gene in said cancer with a Pearson's coefficient of ≥ 0.5 .

The patient can be a mammal, including higher primates, such as humans, and is preferably a human patient.

Disease outcome can be expressed in various forms, including overall patient survival, recurrence-free survival, or distant recurrence-free survival.

In a particular embodiment, the prognosis assumes that the patient receives no further treatment after surgical resection of said breast cancer.

In another embodiment, the expression levels are normalized relative to the expression levels of one or more reference genes, or their expression products, where the reference genes may be selected, for example, from the group consisting of ACTB, GAPD, GUS, RPLPO, and TFRC.

In yet another embodiment, the expression levels are normalized relative to the mean of the expression levels of ACTB, GAPD, GUS, RPLPO, and TFRC.

In a specific embodiment, the quantitative value of said disease outcome is directly proportionate with the value of the variable or variables determined over a continuum.

In further embodiments, the method comprises the determination of the Proliferation Score, and optionally one or both of the Proliferation Group Threshold Score and the Invasion Group Score.

The method of the invention may comprise determination of the expression levels of at least two, or at least three, or at least four, or at least five, or at least six, or at least seven, or at least eight of the individual genes listed in (iv), or their expression products.

In a specific embodiment, the method comprises determination of the expression levels of all individual genes listed in (iv), or their expression products.

The breast cancer may, for example, be lymph node negative and/or ESR1 positive.

The method of the present invention can be performed more than once on the same patient, such as, prior to and following chemotherapy, hormonal therapy and/or radiation therapy.

5 If the patient is determined to have an increased risk of poor disease outcome, it can be treated with chemotherapy, hormonal therapy and/or radiation therapy, following such determination, where chemotherapy includes all chemotherapy approaches used in clinical practice, including adjuvant and neoadjuvant chemotherapy.

10 In a particular embodiment, the chemotherapy comprises the administration of a taxane derivative, e.g. docetaxel or paclitaxel.

In another embodiment, the chemotherapy comprises the administration of an anthracycline derivative, e.g. doxorubicin.

15 In yet another embodiment, the chemotherapy comprises the administration of a topoisomerase inhibitor, e.g. camptothecin, topotecan, irinotecan, 20-S-camptothecin, 9-nitro-camptothecin, 9-amino-camptothecin, or GI147211.

In a further embodiment, the hormonal therapy comprises the administration of TAM.

20 In yet another embodiment, the hormonal therapy comprises the administration of an anti-estrogen drug, which may be an antagonist of estrogen binding to the estrogen receptor, or an inhibitor of estrogen biosynthesis, such as an aromatase inhibitor. Specific representatives of such anti-estrogen drugs include toremifene, anastrozole, and megesterol acetate.

The biological sample assayed by the methods of the present invention can be a sample from a solid tumor, i.e. a tissue sample comprising cancer cells.

25 The tissue may, for example, be fixed, paraffin-embedded, or fresh, or frozen, and can derive from fine needle, core, or other types of biopsy. In particular embodiments, the tissue sample is obtained by fine needle aspiration, bronchial lavage, or transbronchial biopsy.

In a further embodiment, the gene expression levels are determined by quantitative RT-PCR.

30 In yet another embodiment, the expression level of the expression product or products is determined by immunohistochemistry or by proteomics techniques.

In a still further embodiment, the method of the invention further comprises the step of creating a report summarizing the prognosis.

In another aspect, the invention concerns a method for quantitative determination of the likelihood of a beneficial response of an ESR1 positive breast cancer patient to treatment with an anti-estrogen drug, comprising quantitatively determining, in a biological sample comprising cancer cells obtained from said patient, one or more of the following variables:

(i) ESR1 Group Score; and

(ii) the expression levels of the RNA transcripts of one or more of the following individual genes ESR1, SCUBE2, TFRC, and BCL2, or their expression products; wherein

for every unit of increased numerical value of ESR1 Group Score, ESR1, SCUBE2, or BCL2 the patient is identified to have a proportionately increased likelihood of a beneficial response to treatment with an anti-estrogen drug, and for every unit of increased numerical value of TFRC the patient is identified as having a decreased likelihood of beneficial response to treatment with an anti-estrogen drug.

In one embodiment, the ESR1 Group Score, or the expression level of the ESR1 gene or its expression product is determined.

In another embodiment, the expression level of the ESR1 gene or its expression product is determined.

The anti-estrogen drug may, for example, be selected from the group consisting of tamoxifen, toremifene, anastrozole, and megesterol acetate.

In another embodiment, the anti-estrogen drug is TAM.

The treatment predictive method of the present invention may comprising the step of preparing a report for the patient, including a treatment recommendation.

In a further embodiment, anti-estrogen drug treatment without chemotherapy is recommended when ESR1 expression for the patient is higher than the expression value measured in the same test for percentage of ER-positive, node-negative breast cancer patients, and the patient is otherwise known to be in a low risk group.

In another embodiment, the variable is the expression level of the RNA transcript of ESR1, or its expression product.

In a further embodiment, the method includes the step of determining the Recurrence Score for the patient.

In a particular embodiment, after determining the Recurrence Score, a treatment comprising chemotherapy is recommended when ESR1 expression for the patient is non-zero but is lower than the expression value measured in the same test for a particular percentage of ER-positive, node-negative breast cancer patients and the patient is otherwise known to be in a group having a high risk of recurrence.

In yet another aspect, the invention concerns a kit comprising a set of gene specific probes and/or primers for quantifying the expression of one or more of the genes listed in Table 1 by quantitative RT-PCR.

In a particular embodiment, the gene specific probes are selected from the group consisting of the probes listed in Table 7.

In another embodiment, the gene specific primers are selected from the group consisting of the forward and reverse primers listed in Table 7.

In yet another embodiment, the amplicon amplified by the quantitative RT-PCR is selected from the amplicons listed in Table 8.

In further embodiments, the kits of the present invention may comprise one or more reagents for expression of RNA from tumor samples, and/or one or more, where the containers may, for example, comprise pre-fabricated microarrays, a buffers, nucleotide triphosphates, reverse transcriptase, DNA polymerase, RNA polymerase, probes, or primers. The kit may additionally comprise a label or package insert with instructions for use of its components, such as, for example, for use in the prediction or prognosis of breast cancer.

Brief Description of the Figures

Figure 1 shows the absolute increase in proportion of TAM treated patients who are distant disease-free at 10 years as a function of quantitative measurement of ESR1

Figure 2 shows the absolute benefit of tamoxifen (TAM) treatment as a function of ESR1 expression. The horizontal lines through bars represent 95% confidence limits.

Figure 3 shows the absolute benefit of TAM treatment as a function of the quartile into which the expression of ESR1 falls. The horizontal lines through bars represent 95%

confidence limits. The cutpoints defining the quartiles of ESR1 expression, are based on reference-normalized C_T values derived from analysis of the NSABP B-14 patient population.

Detailed Description of the Preferred Embodiment

5 A. Definitions

Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Singleton *et al.*, Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, NY 1994); and Webster's New World™ Medical Dictionary, 2nd
10 Edition, Wiley Publishing Inc., 2003, provide one skilled in the art with a general guide to many of the terms used in the present application. For purposes of the present invention, the following terms are defined below.

The term "beneficial response" means an improvement in any measure of patient status, including those measures ordinarily used in the art, such as overall survival,
15 progression free survival, recurrence-free survival, and distant recurrence-free survival. Recurrence-free survival (RFS) refers to the time (in years) from surgery to the first local, regional, or distant recurrence. Distant recurrence-free survival (DFRS) refers to the time (in years) from surgery to the first anatomically distant recurrence. The calculation of these measures in practice may vary from study to study depending on the definition of events to be
20 either censored or not considered.

The term "microarray" refers to an ordered arrangement of hybridizable array elements, preferably polynucleotide probes, on a substrate.

The term "polynucleotide," when used in singular or plural, generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or
25 modified RNA or DNA. Thus, for instance, polynucleotides as defined herein include, without limitation, single- and double-stranded DNA, DNA including single- and double-stranded regions, single- and double-stranded RNA, and RNA including single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or include single- and double-stranded regions. In
30 addition, the term "polynucleotide" as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of

the molecules of a triple-helical region often is an oligonucleotide. The term “polynucleotide” specifically includes cDNAs. The term includes DNAs (including cDNAs) and RNAs that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are “polynucleotides” as that term is intended
5 herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritiated bases, are included within the term “polynucleotides” as defined herein. In general, the term “polynucleotide” embraces all chemically, enzymatically and/or metabolically modified forms of unmodified polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including simple and complex cells.

10 The term “oligonucleotide” refers to a relatively short polynucleotide, including, without limitation, single-stranded deoxyribonucleotides, single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs. Oligonucleotides, such as single-stranded DNA probe oligonucleotides, are often synthesized by chemical methods, for example using automated oligonucleotide synthesizers that are commercially available.
15 However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms.

The term “gene expression” describes the conversion of the DNA gene sequence information into transcribed RNA (the initial unspliced RNA transcript or the mature mRNA) or the encoded protein product. Gene expression can be monitored by measuring the levels
20 of either the entire RNA or protein products of the gene or subsequences.

The phrase “gene amplification” refers to a process by which multiple copies of a gene or gene fragment are formed in a particular cell or cell line. The duplicated region (a stretch of amplified DNA) is often referred to as “amplicon.” Often, the amount of the messenger RNA (mRNA) produced, *i.e.*, the level of gene expression, also increases in the
25 proportion of the number of copies made of the particular gene expressed.

Prognostic factors are those variables related to the natural history of breast cancer, which influence the recurrence rates and outcome of patients once they have developed breast cancer. Clinical parameters that have been associated with a worse prognosis include, for example, lymph node involvement, increasing tumor size, and high grade tumors. Prognostic
30 factors are frequently used to categorize patients into subgroups with different baseline relapse risks. In contrast, treatment predictive factors are variables related to the likelihood of an individual patient’s beneficial response to a treatment, such as anti-estrogen or chemotherapy, independent of prognosis.

The term “prognosis” is used herein to refer to the likelihood of cancer-attributable death or cancer progression, including recurrence and metastatic spread of a neoplastic disease, such as breast cancer, during the natural history of the disease. Prognostic factors are those variables related to the natural history of a neoplastic diseases, such as breast cancer, which influence the recurrence rates and disease outcome once the patient developed the neoplastic disease, such as breast cancer. In this context, “natural outcome” means outcome in the absence of further treatment. For example, in the case of breast cancer, “natural outcome” means outcome following surgical resection of the tumor, in the absence of further treatment (such as, chemotherapy or radiation treatment). Prognostic factors are frequently used to categorize patients into subgroups with different baseline risks, such as baseline relapse risks.

The term “prediction” is used herein to refer to the likelihood that a patient will respond either favorably or unfavorably to a drug or set of drugs, and also the extent of those responses. Thus, treatment predictive factors are those variables related to the response of an individual patient to a specific treatment, independent of prognosis. The predictive methods of the present invention can be used clinically to make treatment decisions by choosing the most appropriate treatment modalities for any particular patient. The predictive methods of the present invention are valuable tools in predicting if a patient is likely to respond favorably to a treatment regimen, such as anti-estrogen therapy, such as TAM treatment alone or in combination with chemotherapy and/or radiation therapy.

The term “long-term” survival is used herein to refer to survival for at least 3 years, more preferably for at least 8 years, most preferably for at least 10 years following surgery or other treatment.

The term “tumor,” as used herein, refers to all neoplastic cell growth and proliferation, whether malignant or benign, and all pre-cancerous and cancerous cells and tissues.

The terms “cancer” and “cancerous” refer to or describe the physiological condition in mammals that is typically characterized by unregulated cell growth. Examples of cancer include, but are not limited to, breast cancer, ovarian cancer, colon cancer, lung cancer, prostate cancer, hepatocellular cancer, gastric cancer, pancreatic cancer, cervical cancer, liver cancer, bladder cancer, cancer of the urinary tract, thyroid cancer, renal cancer, carcinoma, melanoma, and brain cancer.

The "pathology" of cancer includes all phenomena that compromise the well-being of the patient. This includes, without limitation, abnormal or uncontrollable cell growth, metastasis, interference with the normal functioning of neighboring cells, release of cytokines or other secretory products at abnormal levels, suppression or aggravation of inflammatory or immunological response, neoplasia, premalignancy, malignancy, invasion of surrounding or distant tissues or organs, such as lymph nodes, etc.

In the context of the present invention, reference to "at least one," "at least two," "at least three," "at least four," "at least five," etc. of the genes listed in any particular gene set means any one or any and all combinations of the genes listed.

The term "node negative" cancer, such as "node negative" breast cancer, is used herein to refer to cancer that has not spread to the lymph nodes.

The terms "splicing" and "RNA splicing" are used interchangeably and refer to RNA processing that removes introns and joins exons to produce mature mRNA with continuous coding sequence that moves into the cytoplasm of an eukaryotic cell.

In theory, the term "exon" refers to any segment of an interrupted gene that is represented in the mature RNA product (B. Lewin. *Genes IV* Cell Press, Cambridge Mass. 1990). In theory the term "intron" refers to any segment of DNA that is transcribed but removed from within the transcript by splicing together the exons on either side of it. Operationally, exon sequences occur in the mRNA sequence of a gene as defined by Ref. SEQ ID numbers. Operationally, intron sequences are the intervening sequences within the genomic DNA of a gene, bracketed by exon sequences and having GT and AG splice consensus sequences at their 5' and 3' boundaries.

B. Detailed Description

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, and biochemistry, which are within the skill of the art. Such techniques are explained fully in the literature, such as, "Molecular Cloning: A Laboratory Manual", 2nd edition (Sambrook et al., 1989); "Oligonucleotide Synthesis" (M.J. Gait, ed., 1984); "Animal Cell Culture" (R.I. Freshney, ed., 1987); "Methods in Enzymology" (Academic Press, Inc.); "Handbook of Experimental Immunology", 4th edition (D.M. Weir & C.C. Blackwell, eds., Blackwell Science Inc., 1987); "Gene Transfer Vectors for Mammalian Cells" (J.M. Miller & M.P. Calos, eds., 1987); "Current Protocols in Molecular Biology"

(F.M. Ausubel et al., eds., 1987); and "PCR: The Polymerase Chain Reaction", (Mullis et al., eds., 1994).

B.1. General Description of the Invention

Over the past two years Genomic Health, Inc and collaborators (Esteban et al., *Proc Am Soc Clin Oncol* 22: page 850, 2003 (abstract 3416); Cobleigh et al. *Soc Clin Oncol* 22: page 850, 2003 (abstract 3415); Soule et al., *Proc Am Soc Clin Oncol* 22: page 862, 2003 (abstract 3466)) reported several exploratory clinical studies of gene expression in early breast cancer, aimed at finding a molecular signature for recurrence risk. These studies used quantitative RT-PCR to test 250 candidate gene markers in frozen, paraffin-embedded (FPE) tissue specimens having linked clinical records. Analysis across all three studies was performed in order to examine whether genes could be identified which were consistently related to the risk of recurrence across a diverse group of patients. Based on these univariate results, multi-gene models were designed and analyzed across the three studies. A single multi-gene assay, consisting of 16 cancer-related genes and 5 reference genes, was developed to be tested prospectively in clinical validation studies. An algorithm called Recurrence Score (RS) was generated, which utilizes the measurements of these 21 mRNA species and reads out recurrence risk on a 100 point scale.

To test the clinical validity of this Recurrence Score test and algorithm, a blinded clinical trial with prospectively identified endpoints was carried out. This validation trial focused on patients treated with TAM alone in the randomized and registration arms of the National Surgical Adjuvant Breast and Bowel Project (NSABP) Study B-14 clinical trial population (Fisher B, Costantino JP, Redmond CK, et al: Endometrial cancer in TAM-treated breast cancer patients: Findings from the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14. *J Natl Cancer Inst* 86:527-537 (1994)). Genomic Health, Inc. and the NSABP carried out the 21 gene RT-PCR assay on 668 breast cancer tissue specimens derived from these patients and calculated a Recurrence Score for each patient.

Pre-specified cut-off points of Recurrence Score classified patients into one of three categories: low risk, intermediate risk, and high risk of distant disease recurrence. The proportion of the 668 patients categorized as low, intermediate, and high risk by the RT-PCR assay were 51%, 23%, and 27%, respectively. The Kaplan-Meier estimates and 95% confidence intervals for the rates of distant recurrence at 10 years were 6.8% (4.0%, 9.6%), 14.3% (8.3%, 20.3%) 30.5% (23.6%, 37.4%), respectively, for the low, intermediate, and

high risk groups; the rate for the low risk group was significantly lower than the rate for the high risk group ($p < 0.001$). In a multivariate Cox model relating distant recurrence to Recurrence Score, age, and tumor size, Recurrence Score provides significant ($p < 0.001$) predictive power that goes beyond age and tumor size. This study validated the Recurrence Score as a powerful predictor of distant recurrence in patients without involved nodes who have tumors that are ESR1 positive and treated with TAM (Paik et al. Breast Cancer Research and Treatment 82, Supplement 1: page S10, 2003 (Abstract 16)).

The invention disclosed herein derives, in part, from study of patients in the placebo arm of the NSABP B-14 clinical study (B-14) and, in part, from comparison of patients in the B-14 placebo arm to TAM-treated patients with patients in the randomized and registration arms of NSABP Study B-14. Breast cancer tissue derived from placebo-treated patients was quantitatively analyzed, using a RT-PCR assay to quantify the expression of sixteen cancer-related genes and five reference genes.

The quantitative gene expression analysis resulted in the identification of molecular indicators of prognosis. Based on analysis of the relationship between gene expression and distant recurrence-free survival in the placebo arm of the NSABP B-14 trial, a set of genes has been identified, the expression levels of which are indicative of outcome if no further treatment is provided to the patient. Outcome may be manifest in various measurements including survival, recurrence-free survival and distant recurrence-free survival, all of which are within the scope of the invention.

The prognostic genes and gene groups identified may be used singly or in particular combinations to predict outcome likelihood for particular patients. Prognostic indicators include, specifically, the proliferation group (BIRC5 + MKI + MYBL2 + CCNB1 + STK6), the invasion group (CTSL2 + MMP11), and one or more of the individual genes: CCNB1, BIRC5, MYBL2, PGR, STK6, MKI, GSTM1, GAPD, RPLP0, and MMP11.

In another aspect, the gene expression analysis disclosed herein resulted in the identification of molecular indicators of beneficial response to anti-estrogen drugs, such as TAM, based on analysis of the relationship between gene expression and distant recurrence-free survival in untreated patients from the placebo arm of B-14 as well as TAM treated patients from both registration and randomization arms of B-14.

Based on interaction analysis of the relationship between gene expression and distant recurrence-free survival in the combined placebo and treatment cohorts, a set of genes and gene groups has been identified, the expression levels of which are indicative of beneficial response to treatment TAM. These genes and gene groups may be used singly or in particular combinations to predict likelihood of beneficial response to treatment with TAM, or another anti-estrogen drug, for individual patients. Specifically, these genes/gene groups are: the ESR1 group (ESR1 + PGR + BCL2 + SCUBE2) and one or more of the individual genes ESR1, SCUBE2, TFRC, and BCL2.

A significant finding of the invention is that quantitative levels of ESR1 relate to likelihood of TAM benefit as a continuous variable across a 14 point expression scale. Thus, for an individual patient it is possible to provide a quantitative estimate of likelihood of benefit for this therapeutic agent, with higher ESR1 gene expression levels correlating with a greater chance of response. This information can be utilized in several ways. It provides a more refined assessment of the probability of a beneficial response to TAM treatment, and other anti-estrogen therapies, than has been available previously. TAM has significant side effects, including development of uterine cancer, deep vein thrombosis, pulmonary embolism, and cataracts (Physicians Desk Reference 2002). Similarly, other anti-estrogen drugs, such as, toremifene (Fareston®, Orion, Corp.), anastrozole (Arimidex®, AstraZeneca Pharmaceuticals), and megesterol acetate, have serious side-effects. As a result of the present invention, patients and their oncologists can now use the ESR1 score to assess risk versus benefit when deciding whether TAM treatment, or other anti-estrogen therapy, is appropriate.

While expression of ESR1 (principally as determined by immunochemistry measurement at the protein level) is routinely used in clinical practice to determine whether a patient should be treated with TAM, based on “ESR1 positive” or “ESR1 negative” status, the findings underlying the present invention relate to patients who are already defined as “ESR1 positive” by the conventional criteria. According to the present invention, it is possible to determine the likelihood of a beneficial response to TAM treatment, or treatment with other anti-estrogen drugs, among this group of patients.

ESR1 levels can be used in conjunction with the Recurrence Score (discussed below) to determine whether individual patients should be prescribed TAM (or another anti-estrogen drug) alone, or TAM (or another anti-estrogen drug) plus chemotherapy.

The invention additionally allows the design of a particular test, an example of which is given in the Example below, with precise ESR1 expression cut points that predict a high, intermediate or low level of benefit from TAM treatment, or treatment with other anti-estrogen drugs.

5 In various embodiments of the inventions, various technological approaches are available for determination of expression levels of the disclosed genes, including, without limitation, RT-PCR, microarrays, serial analysis of gene expression (SAGE) and Gene Expression Analysis by Massively Parallel Signature Sequencing (MPSS), which will be discussed in detail below. In particular embodiments, the expression level of each gene may
10 be determined in relation to various features of the expression products of the gene including exons, introns, protein epitopes and protein activity.

B.2 Gene Expression Profiling

In general, methods of gene expression profiling can be divided into two large groups: methods based on hybridization analysis of polynucleotides, and methods based on
15 sequencing of polynucleotides. The most commonly used methods known in the art for the quantification of mRNA expression in a sample include northern blotting and *in situ* hybridization (Parker & Barnes, *Methods in Molecular Biology* 106:247-283 (1999)); RNase protection assays (Hod, *Biotechniques* 13:852-854 (1992)); and reverse transcription polymerase chain reaction (RT-PCR) (Weis *et al.*, *Trends in Genetics* 8:263-264 (1992)).
20 Alternatively, antibodies may be employed that can recognize specific duplexes, including DNA duplexes, RNA duplexes, and DNA-RNA hybrid duplexes or DNA-protein duplexes. Representative methods for sequencing-based gene expression analysis include Serial Analysis of Gene Expression (SAGE), and gene expression analysis by massively parallel signature sequencing (MPSS).

25 Two biological processes commonly involved in tumorigenesis include gene amplification and DNA methylation. Both processes result in the abnormal expression of genes important in tumor formation or progression. Methods that monitor gene amplification and DNA methylation can therefore be considered surrogate methods for gene expression profiling.

30 Gene amplification is a common alteration in many cancers that can lead to elevated expression of cellular oncogenes (Meltzer, P. *et al.*, *Cancer Genet Cytogenet.* 19:93 (1986). In breast cancer, there is good correlation between ERBB2 gene amplification and ERBB2

overexpression (Nagai, M.A. et al., Cancer Biother 8:29 (1993), Savinainen, K.J. et al., Am. J. Pathol. 160:339 (2002)). Amplification of the ERBB2 gene, leading to its overexpression, correlates with poor prognosis (Press, M.F. et al., J. Clin. Oncol. 15:2894 (1997), Slamon, D.J. et al., Science 244:707 (1989)) and is predictive for response to anti-HER2 therapy in combination with standard chemotherapy (Seidman, A.D. et al., J. Clin. Oncol. 19:1866 (2001)).

DNA methylation has also been shown to be a common alteration in cancer leading to elevated or decreased expression of a broad spectrum of genes (Jones, P.A. Cancer Res. 65:2463 (1996)). In general, hypomethylation of CpG islands in the promoter regions and regulatory elements results in increased gene expression, including many oncogenes (Hanada, M., et al., Blood 82:1820 (1993), Feinberg, A.P. and Vogelstein, B. Nature 301:89 (1983)). Because DNA methylation correlates with the level of specific gene expression in many cancers, it serves as a useful surrogate to expression profiling of tumors (Toyota, M. et al., Blood 97: 2823 (2001), Adorjan, P. et al. Nucl. Acids. Res. 10:e21 (2002)).

a. Reverse Transcriptase PCR (RT-PCR)

Of the techniques listed above, the most sensitive and most flexible quantitative method is RT-PCR, which can be used to compare mRNA levels in different sample populations, in normal and tumor tissues, with or without drug treatment, to characterize patterns of gene expression, to discriminate between closely related mRNAs, and to analyze RNA structure.

The first step is the isolation of mRNA from a target sample. The starting material is typically total RNA isolated from human tumors or tumor cell lines, and corresponding normal tissues or cell lines, respectively. Thus RNA can be isolated from a variety of primary tumors, including breast, lung, colon, prostate, brain, liver, kidney, pancreas, spleen, thymus, testis, ovary, uterus, etc., tumor, or tumor cell lines, with pooled DNA from healthy donors. If the source of mRNA is a primary tumor, mRNA can be extracted, for example, from frozen or archived paraffin-embedded and fixed (e.g. formalin-fixed) tissue samples.

General methods for mRNA extraction are well known in the art and are disclosed in standard textbooks of molecular biology, including Ausubel *et al.*, Current Protocols of Molecular Biology, John Wiley and Sons (1997). Methods for RNA extraction from paraffin embedded tissues are disclosed, for example, in Rupp and Locker, *Lab Invest.* 56:A (1987), and De Andrés *et al.*, *BioTechniques* 18:42044 (1995). In particular, RNA isolation can be

performed using purification kit, buffer set and protease from commercial manufacturers, such as Qiagen, according to the manufacturer's instructions. For example, total RNA from cells in culture can be isolated using Qiagen RNeasy mini-columns. Other commercially available RNA isolation kits include MasterPure™ Complete DNA and RNA Purification Kit (EPICENTRE®, Madison, WI), and Paraffin Block RNA Isolation Kit (Ambion, Inc.). Total RNA from tissue samples can be isolated using RNA Stat-60 (Tel-Test). RNA prepared from tumor can be isolated, for example, by cesium chloride density gradient centrifugation.

As RNA cannot serve as a template for PCR, the first step in gene expression profiling by RT-PCR is the reverse transcription of the RNA template into cDNA, followed by its exponential amplification in a PCR reaction. The two most commonly used reverse transcriptases are avilo myeloblastosis virus reverse transcriptase (AMV-RT) and Moloney murine leukemia virus reverse transcriptase (MMLV-RT). The reverse transcription step is typically primed using specific primers, random hexamers, or oligo-dT primers, depending on the circumstances and the goal of expression profiling. For example, extracted RNA can be reverse-transcribed using a GeneAmp RNA PCR kit (Perkin Elmer, CA, USA), following the manufacturer's instructions. The derived cDNA can then be used as a template in the subsequent PCR reaction.

Although the PCR step can use a variety of thermostable DNA-dependent DNA polymerases, it typically employs the Taq DNA polymerase, which has a 5'-3' nuclease activity but lacks a 3'-5' proofreading endonuclease activity. Thus, TaqMan® PCR typically utilizes the 5'-nuclease activity of Taq or Tth polymerase to hydrolyze a hybridization probe bound to its target amplicon, but any enzyme with equivalent 5' nuclease activity can be used. Two oligonucleotide primers are used to generate an amplicon typical of a PCR reaction. A third oligonucleotide, or probe, is designed to detect nucleotide sequence located between the two PCR primers. The probe is non-extendible by Taq DNA polymerase enzyme, and is labeled with a reporter fluorescent dye and a quencher fluorescent dye. Any laser-induced emission from the reporter dye is quenched by the quenching dye when the two dyes are located close together as they are on the probe. During the amplification reaction, the Taq DNA polymerase enzyme cleaves the probe in a template-dependent manner. The resultant probe fragments disassociate in solution, and signal from the released reporter dye is free from the quenching effect of the second fluorophore. One molecule of reporter dye is

liberated for each new molecule synthesized, and detection of the unquenched reporter dye provides the basis for quantitative interpretation of the data.

TaqMan® RT-PCR can be performed using commercially available equipment, such as, for example, ABI PRISM 7700™ Sequence Detection System™ (Perkin-Elmer-Applied Biosystems, Foster City, CA, USA), or Lightcycler (Roche Molecular Biochemicals, Mannheim, Germany). In a preferred embodiment, the 5' nuclease procedure is run on a real-time quantitative PCR device such as the ABI PRISM 7700™ Sequence Detection System™. The system consists of a thermocycler, laser, charge-coupled device (CCD), camera and computer. The system amplifies samples in a 96-well format on a thermocycler. During amplification, laser-induced fluorescent signal is detected at the CCD. The system includes software for running the instrument and for analyzing the data.

5'-Nuclease assay data are initially expressed as C_T , or the threshold cycle. As discussed above, fluorescence values are recorded during every cycle and represent the amount of product amplified to that point in the amplification reaction. The point when the fluorescent signal is first recorded as statistically significant is the threshold cycle (C_T).

To minimize errors and the effect of sample-to-sample variation, RT-PCR is usually performed using one or more reference genes as internal standards. The ideal internal standard is expressed at a constant level among different tissues, and is unaffected by the experimental treatment. RNAs most frequently used to normalize patterns of gene expression are mRNAs for the housekeeping genes glyceraldehyde-3-phosphate-dehydrogenase (GAPD) and β -actin (ACTB).

A more recent variation of the RT-PCR technique is the real time quantitative PCR, which measures PCR product accumulation through a dual-labeled fluorogenic probe (i.e., TaqMan® probe). Real time PCR is compatible both with quantitative competitive PCR, where internal competitor for each target sequence is used for normalization, and with quantitative comparative PCR using a normalization gene contained within the sample, or a housekeeping gene for RT-PCR. For further details see, e.g. Held *et al.*, *Genome Research* 6:986-994 (1996).

The steps of a representative protocol for profiling gene expression using fixed, paraffin-embedded tissues as the RNA source, including mRNA isolation, purification, primer extension and amplification are given in various published journal articles {for

example: T.E. Godfrey et al. J. Molec. Diagnostics 2: 84-91 (2000); K. Specht et al., Am. J. Pathol. 158: 419-29 (2001); Cronin et al., Am J Pathol 164:35-42 (2004)}. Briefly, a representative process starts with cutting about 10 μ m thick sections of paraffin-embedded tumor tissue samples. The RNA is then extracted, and protein and DNA are removed. After
5 analysis of the RNA concentration, RNA repair and/or amplification steps may be included, if necessary, and RNA is reverse transcribed using gene specific promoters followed by RT-PCR.

b. Microarrays

Differential gene expression can also be identified, or confirmed using the microarray
10 technique. Thus, the expression profile of breast cancer-associated genes can be measured in either fresh or paraffin-embedded tumor tissue, using microarray technology. In this method, polynucleotide sequences of interest (including cDNAs and oligonucleotides) are plated, or arrayed, on a microchip substrate. The arrayed sequences are then hybridized with specific
15 DNA probes from cells or tissues of interest. Just as in the RT-PCR method, the source of mRNA typically is total RNA isolated from human tumors or tumor cell lines, and corresponding normal tissues or cell lines. Thus RNA can be isolated from a variety of primary tumors or tumor cell lines. If the source of mRNA is a primary tumor, mRNA can be extracted, for example, from frozen or archived paraffin-embedded and fixed (e.g. formalin-fixed) tissue samples, which are routinely prepared and preserved in everyday clinical
20 practice.

In a specific embodiment of the microarray technique, PCR amplified inserts of cDNA clones are applied to a substrate in a dense array. Preferably at least 10,000 nucleotide sequences are applied to the substrate. The microarrayed genes, immobilized on the microchip at 10,000 elements each, are suitable for hybridization under stringent conditions.
25 Fluorescently labeled cDNA probes may be generated through incorporation of fluorescent nucleotides by reverse transcription of RNA extracted from tissues of interest. Labeled cDNA probes applied to the chip hybridize with specificity to each spot of DNA on the array. After stringent washing to remove non-specifically bound probes, the chip is scanned by confocal laser microscopy or by another detection method, such as a CCD camera.
30 Quantitation of hybridization of each arrayed element allows for assessment of corresponding mRNA abundance. With dual color fluorescence, separately labeled cDNA probes generated from two sources of RNA are hybridized pairwise to the array. The relative abundance of the

transcripts from the two sources corresponding to each specified gene is thus determined simultaneously. The miniaturized scale of the hybridization affords a convenient and rapid evaluation of the expression pattern for large numbers of genes. Such methods have been shown to have the sensitivity required to detect rare transcripts, which are expressed at a few
5 copies per cell, and to reproducibly detect at least approximately two-fold differences in the expression levels (Schena *et al.*, *Proc. Natl. Acad. Sci. USA* 93(2):106-149 (1996)). Microarray analysis can be performed by commercially available equipment, following manufacturer's protocols, such as by using the Affymetrix GenChip technology, or Incyte's microarray technology.

10 The development of microarray methods for large-scale analysis of gene expression makes it possible to search systematically for molecular markers of cancer classification and outcome prediction in a variety of tumor types.

c. Serial Analysis of Gene Expression (SAGE)

Serial analysis of gene expression (SAGE) is a method that allows the simultaneous
15 and quantitative analysis of a large number of gene transcripts, without the need of providing an individual hybridization probe for each transcript. First, a short sequence tag (about 10-14 bp) is generated that contains sufficient information to uniquely identify a transcript, provided that the tag is obtained from a unique position within each transcript. Then, many transcripts are linked together to form long serial molecules, that can be sequenced, revealing
20 the identity of the multiple tags simultaneously. The expression pattern of any population of transcripts can be quantitatively evaluated by determining the abundance of individual tags, and identifying the gene corresponding to each tag. For more details see, e.g. Velculescu *et al.*, *Science* 270:484-487 (1995); and Velculescu *et al.*, *Cell* 88:243-51 (1997).

d. Gene Expression Analysis by Massively Parallel Signature Sequencing (MPSS)
25

This method, described by Brenner *et al.*, *Nature Biotechnology* 18:630-634 (2000), is a sequencing approach that combines non-gel-based signature sequencing with *in vitro* cloning of millions of templates on separate 5 µm diameter microbeads. First, a microbead library of DNA templates is constructed by *in vitro* cloning. This is followed by the
30 assembly of a planar array of the template-containing microbeads in a flow cell at a high density (typically greater than 3×10^6 microbeads/cm²). The free ends of the cloned templates on each microbead are analyzed simultaneously, using a fluorescence-based

signature sequencing method that does not require DNA fragment separation. This method has been shown to simultaneously and accurately provide, in a single operation, hundreds of thousands of gene signature sequences from a yeast cDNA library.

e. General Description of the mRNA Isolation, Purification and Amplification

5 The steps of a representative protocol for profiling gene expression using fixed, paraffin-embedded tissues as the RNA source, including mRNA isolation, purification, primer extension and amplification are provided in various published journal articles (for example: T.E. Godfrey *et al.*, *J. Molec. Diagnostics* 2: 84-91 [2000]; K. Specht *et al.*, *Am. J. Pathol.* 158: 419-29 [2001]). Briefly, a representative process starts with cutting about 10
10 μ m thick sections of paraffin-embedded tumor tissue samples. The RNA is then extracted, and protein and DNA are removed. After analysis of the RNA concentration, RNA repair and/or amplification steps may be included, if necessary, and RNA is reverse transcribed using gene specific promoters followed by RT-PCR. Finally, the data are analyzed to identify the best treatment option(s) available to the patient on the basis of the characteristic
15 gene expression pattern identified in the tumor sample examined, dependent on the predicted likelihood of cancer recurrence.

f. Breast Cancer Gene Set, Assayed Gene Subsequences, and Clinical Application of Gene Expression Data

An important aspect of the present invention is to use the measured expression of
20 certain genes by breast cancer tissue to provide prognostic or predictive information. For this purpose it is necessary to correct for (normalize away) both differences in the amount of RNA assayed and variability in the quality of the RNA used. Therefore, the assay typically measures and incorporates the expression of certain normalizing genes, including well known housekeeping genes, such as β -actin, GAPD, GUS, RPLO, and TFRC, as shown in the
25 Example below. Alternatively, normalization can be based on the mean or median signal (C_T) of all of the assayed genes or a large subset thereof (global normalization approach). Below, unless noted otherwise, gene expression means normalized expression.

g. Design of Intron-Based PCR Primers and Probes

According to one aspect of the present invention, PCR primers and probes are
30 designed based upon intron sequences present in the gene to be amplified. Accordingly, the first step in the primer/probe design is the delineation of intron sequences within the genes. This can be done by publicly available software, such as the DNA BLAT software developed

by Kent, W.J., *Genome Res.* 12(4):656-64 (2002), or by the BLAST software including its variations. Subsequent steps follow well established methods of PCR primer and probe design.

In order to avoid non-specific signals, it is important to mask repetitive sequences within the introns when designing the primers and probes. This can be easily accomplished by using the Repeat Masker program available on-line through the Baylor College of Medicine, which screens DNA sequences against a library of repetitive elements and returns a query sequence in which the repetitive elements are masked. The masked intron sequences can then be used to design primer and probe sequences using any commercially or otherwise publicly available primer/probe design packages, such as Primer Express (Applied Biosystems); MGB assay-by-design (Applied Biosystems); Primer3 (Steve Rozen and Helen J. Skaletsky (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386).

The most important factors considered in PCR primer design include primer length, melting temperature (T_m), and G/C content, specificity, complementary primer sequences, and 3'-end sequence. In general, optimal PCR primers are generally 17-30 bases in length, and contain about 20-80%, such as, for example, about 50-60% G+C bases. T_m 's between 50 and 80 °C, e.g. about 50 to 70 °C are typically preferred.

For further guidelines for PCR primer and probe design see, e.g. Dieffenbach, C.W. *et al.*, "General Concepts for PCR Primer Design" in: *PCR Primer, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, 1995, pp. 133-155; Innis and Gelfand, "Optimization of PCRs" in: *PCR Protocols, A Guide to Methods and Applications*, CRC Press, London, 1994, pp. 5-11; and Plasterer, T.N. Primerselect: Primer and probe design. *Methods Mol. Biol.* 70:520-527 (1997), the entire disclosures of which are hereby expressly incorporated by reference.

B.3 Algorithms and Statistical Methods

The present invention takes advantage of certain algorithms and statistical methods, which are described in copending application Serial No. 10/883,303.

When quantitative RT-PCR (qRT-PCR) is used to measure mRNA levels, mRNA amounts are expressed in C_T (threshold cycle) units (Held *et al.*, *Genome Research* 6:986-994

(1996)). The averaged sum of reference mRNA C_T s is set as a fixed number such as zero, and each measured test mRNA C_T is given relative to this fixed point. For example, if, for a certain patient tumor specimen the average of C_T s of the 5 reference genes found to be 31 and C_T of the test gene X is found to be 35, the reported value for gene X is -4 (i.e. 31-35).

5 As a first step following the quantitative determination of mRNA levels, the genes identified in the tumor specimen and known to be associated with the molecular pathology of cancer are grouped into subsets. Thus, genes known to be associated with proliferation will constitute the "Proliferation Group" (axis, or subset). Genes known to be associated with invasion of cancer will constitute the "Invasion Group" (axis, or subset). Genes associated
10 with key growth factor receptor pathway(s) will constitute the "Growth Factor Group" (axis, or subset), also referred to as GRB7 Group. Genes known to be involved with activating or signaling through the estrogen receptor (ESR1) will constitute the "Estrogen Receptor (ER) Group" (axis, or subset), and so on. This list of subsets is, of course, not limiting. The subsets (axes) created will depend on the particular cancer, i.e. breast, prostate, pancreatic,
15 lung, etc. cancer. In general, genes the expression of which is known to correlate with each other, or which are known to be involved in the same pathway are grouped together.

In the next step, the measured tumor level of each mRNA in a subset is multiplied by a coefficient reflecting its relative intra-set contribution to the risk of cancer recurrence and this product is added to the other products between mRNA levels in the subset and their
20 coefficients, to yield a term, e.g. a proliferation term, an invasion term, a growth factor term, etc. For example, in the case of lymph node-negative invasive breast cancer the growth factor term is $(0.45 \text{ to } 1.35) \times \text{GRB7} + (0.05 \text{ to } 0.15) \times \text{ErbB2}$, such as, for example $0.9 \times \text{GRB7} + 0.1 \times \text{ERBB2}$ (see Example below).

The contribution of each term to the overall recurrence score is weighted by use of a
25 coefficient. For example, in the case of lymph node-negative invasive breast cancer the coefficient of the growth factor term can be between 0.23 and 0.70.

Additionally, for some terms, such as the growth factor and proliferation terms, a further step is performed. If the relationship between the term and the risk of recurrence is non-linear, a non-linear functional transform of the term, such as a threshold is used. Thus, in
30 lymph node-negative invasive breast cancer, when the growth factor term is found at <8 the value is fixed at 8.

The sum of the terms obtained provides the recurrence score (RS), which predicts the likelihood of cancer recurrence in the normal course of the disease.

The RS scale generated by the algorithm of the present invention can be adjusted in various ways. Thus, the range could be selected such that the scale run from 0 to 10, 0 to 50, or 0 to 100, for example.

For example, in the particular scaling approach described in the Example below, scaled recurrence score is calculated on a scale of 0 to 100. For convenience, 10 C_T units are added to each measured C_T value, and unscaled RS is calculated as described before. Equations for calculating unscaled RS and scaled RS are provided in the following Example.

In calculating the Recurrence Score, or any variable used to calculate the Recurrence Score, any gene can be substituted by another gene that coexpresses in a set of at least 30 different patient specimens of that tumor type (such as breast cancer) with the first gene in the particular cancer tested with a Pearson's coefficient of ≥ 0.5 . Similarly, any individual gene, or gene within a gene group (subset) included in the prognostic and predictive methods of the present invention can be substituted by another gene that coexpresses with the first gene in the particular cancer tested with a Pearson's coefficient of ≥ 0.5 .

B.4 Anti-Estrogen Drug Treatment

Estrogen is known to promote the growth of some cancers, such as breast cancers, especially those that express the estrogen receptor (ESR1). Several therapies have been developed to block the effect of estrogen, or to lower estrogen levels in such patient, especially in ESR1 positive breast cancer patients.

Anti-estrogen drugs can be generally classified as either antagonists of estrogen binding to the estrogen receptor, or are inhibitors of estrogen biosynthesis, such as aromatase inhibitors.

The most commonly used anti-estrogen drug is TAM, which belong to the class of estrogen binding to the estrogen receptor, is typically taken orally, once a day for five years following surgical resection of breast cancer, and/or chemo- or radiation therapy. Clinical studies have shown that the use of TAM as an adjuvant therapy after surgery reduces the risk of cancer recurrence, however, the response of ESR1 positive patients to this treatment varies, and there have been no clear predictors of responsiveness available.

Other anti-estrogen drugs include raloxifene, which, like TAM, blocks the effect of estrogen on breast tissue and breast cancer; and toremifene citrate, which is closely related to TAM, and may be an option for post menopausal women with metastatic breast cancer.

5 Anastrozole, an aromatase inhibitor, acts by preventing estrogen from activating its receptor, blocking an enzyme needed for production of estrogen. Anastrozole is currently an option for women whose advanced breast cancer continues to grow during or after TAM treatment.

Megesterol acetate is typically used for hormonal treatment of advanced breast cancer, usually for women whose cancers fail to respond to TAM.

10 All treatments are limited by serious side effects, and by the lack of reliable predictors of patient response, which would enable a physician to make an intelligent risk-benefit analysis before recommending a particular treatment.

B.5 Cancer Chemotherapy

Chemotherapeutic agents used in cancer treatment can be divided into several groups, depending on their mechanism of action. Some chemotherapeutic agents directly damage DNA and RNA. By disrupting replication of the DNA such chemotherapeutics either completely halt replication, or result in the production of nonsense DNA or RNA. This category includes, for example, cisplatin (Platinol®), daunorubicin (Cerubidine®), doxorubicin (Adriamycin®), and etoposide (VePesid®). Another group of cancer
20 chemotherapeutic agents interfere with the formation of nucleotides or deoxyribonucleotides, so that RNA synthesis and cell replication is blocked. Examples of drugs in this class include methotrexate (Abitrexate®), mercaptopurine (Purinethol®), fluorouracil (Adrucil®), and hydroxyurea (Hydrea®). A third class of chemotherapeutic agents effects the synthesis or breakdown of mitotic spindles, and, as a result, interrupt cell division. Examples of drugs in
25 this class include Vinblastine (Velban®), Vincristine (Oncovin®) and taxenes, such as, Paclitaxel (Taxol®), and Docetaxel (Taxotere®) Docetaxel is currently approved in the United States to treat patients with locally advanced or metastatic breast cancer after failure of prior chemotherapy, and patients with locally advanced or metastatic non-small cell lung cancer after failure of prior platinum-based chemotherapy. The prediction of patient response to all
30 of these, and other chemotherapeutic agents is specifically within the scope of the present invention.

In a specific embodiment, chemotherapy includes treatment with a taxane derivative. Taxanes include, without limitation, paclitaxel (Taxol®) and docetaxel (Taxotere®), which are widely used in the treatment of cancer. As discussed above, taxanes affect cell structures called microtubules, which play an important role in cell functions. In normal cell growth, microtubules are formed when a cell starts dividing. Once the cell stops dividing, the microtubules are broken down or destroyed. Taxanes stop the microtubules from breaking down; which blocks cancer cell division.

In another specific embodiment, chemotherapy includes treatment with an anthracycline derivative, such as, for example, doxorubicin, daunorubicin, and aclacinomycin.

In a further specific embodiment, chemotherapy includes treatment with a topoisomerase inhibitor, such as, for example, camptothecin, topotecan, irinotecan, 20-S-camptothecin, 9-nitro-camptothecin, 9-amino-camptothecin, or GI147211.

Treatment with any combination of these and other chemotherapeutic drugs is specifically contemplated.

Most patients receive chemotherapy immediately following surgical removal of tumor. This approach is commonly referred to as adjuvant therapy. However, chemotherapy can be administered also before surgery, as so called neoadjuvant treatment. Although the use of neo-adjuvant chemotherapy originates from the treatment of advanced and inoperable breast cancer, it has gained acceptance in the treatment of other types of cancers as well. The efficacy of neoadjuvant chemotherapy has been tested in several clinical trials. In the multi-center National Surgical Adjuvant Breast and Bowel Project B-18 (NSAB B-18) trial (Fisher et al., *J. Clin. Oncology* 15:2002-2004 (1997); Fisher et al., *J. Clin. Oncology* 16:2672-2685 (1998)) neoadjuvant therapy was performed with a combination of adriamycin and cyclophosphamide ("AC regimen"). In another clinical trial, neoadjuvant therapy was administered using a combination of 5-fluorouracil, epirubicin and cyclophosphamide ("FEC regimen") (van Der Hage et al., *J. Clin. Oncol.* 19:4224-4237 (2001)). Newer clinical trials have also used taxane-containing neoadjuvant treatment regimens. See, e.g. Holmes et al., *J. Natl. Cancer Inst.* 83:1797-1805 (1991) and Moliterni et al., *Seminars in Oncology*, 24:S17-10-S-17-14 (1999). For further information about neoadjuvant chemotherapy for breast cancer see, Cleator et al., *Endocrine-Related Cancer* 9:183-195 (2002).

-28-

B.6 Kits of the Invention

The materials for use in the methods of the present invention are suited for preparation of kits produced in accordance with well known procedures. The invention thus provides kits comprising agents, which may include gene-specific or gene-selective probes
5 and/or primers, for quantitating the expression of the disclosed genes for predicting prognostic outcome or response to treatment. Such kits may optionally contain reagents for the extraction of RNA from tumor samples, in particular fixed paraffin-embedded tissue samples and/or reagents for RNA amplification. In addition, the kits may optionally comprise the reagent(s) with an identifying description or label or instructions relating to
10 their use in the methods of the present invention. The kits may comprise containers (including microtiter plates suitable for use in an automated implementation of the method), each with one or more of the various reagents (typically in concentrated form) utilized in the methods, including, for example, pre-fabricated microarrays, buffers, the appropriate nucleotide triphosphates (e.g., dATP, dCTP, dGTP and dTTP; or rATP, rCTP, rGTP and
15 UTP), reverse transcriptase, DNA polymerase, RNA polymerase, and one or more probes and primers of the present invention (e.g., appropriate length poly(T) or random primers linked to a promoter reactive with the RNA polymerase). Mathematical algorithms used to estimate or quantify prognostic or predictive information are also properly potential components of kits.

The methods provided by the present invention may also be automated in whole or in
20 part.

All aspects of the present invention may also be practiced such that a limited number of additional genes that are co-expressed with the disclosed genes, for example as evidenced by high Pearson correlation coefficients, are included in a prognostic or predictive test in addition to and/or in place of disclosed genes.

25 Having described the invention, the same will be more readily understood through reference to the following Example, which is provided by way of illustration, and is not intended to limit the invention in any way.

Example

30 A study of the Relationship between Gene Expression and Prognosis and Likelihood of Beneficial Response to Tamoxifen in Early Breast Cancer Patients

Methods

This study employs tissue and data from NSABP Study B-14: “A Clinical Trial to Assess Tamoxifen in Patients with Primary Breast Cancer and Negative Axillary Nodes Whose Tumors are Positive for Estrogen Receptors.” The results of this trial were reported by Fisher B, Costantino JP, Redmond CK, *et al*: Endometrial cancer in TAM-treated breast cancer patients: Findings from the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14. *J Natl Cancer Inst* 86:527-537 (1994).

1.1 Fixed paraffin-embedded breast tumor tissue samples from up to 450 patients, who were treated at study entry with placebo alone in the B-14 study, were analyzed. For each evaluable patient, the expression of 16 cancer-related genes and 5 reference genes was quantitatively assessed by RT-PCR. The relationship between disease recurrence and (a) recurrence score, (b) expression of genes in particular gene groups, or (c) expression of individual genes was evaluated.

1.2 Inclusion Criteria

1.2.1 Enrolled in NSABP Study B-14: “A Clinical Trial to Assess Tamoxifen in Patients with Primary Breast Cancer and Negative Axillary Nodes Whose Tumors are Positive for Estrogen Receptors.”

1.2.2 Randomization to placebo or to TAM in the placebo-controlled part of the study.

1.2.3 Clinically eligible with follow-up

1.3 Exclusion Criteria

1.3.1 No tumor block available from initial diagnosis in the NSABP archive.

1.3.2 No tumor or very little tumor (Group 1) in block as assessed by examination of the H&E slide.

1.3.3 Insufficient RNA (<275 ng) for RT-PCR analysis.

1.3.4 Average non-normalized C_T for the 5 reference genes >35.

1.4 Gene Panel

-30-

1.4.1 Analysis of 16 cancer-related genes and 5 reference genes listed in Table 1 was carried out using quantitative RT-PCR.

1.4.2 Patient Survival.

Distant recurrence-free survival (DRFS) is based on the time (in years) from surgery to first distant recurrence. Contralateral disease, other second primary cancers, and deaths prior to distant recurrence will be considered censoring events. For the primary analysis, ipsilateral breast recurrence, local chest wall recurrence and regional recurrence is ignored, i.e., not considered either as an event or a censoring event.

1.4.3 Gene Expression.

Expression levels of 21 genes listed in Table 1 were reported as values from the GHI assay. Gene expression values were normalized relative to the mean of the reference genes. For each cancer-related gene, cycle threshold (C_T) measurements were obtained by RT-PCR, and normalized relative to a set of five reference genes listed in Table 1. The reference genes are known to be relatively invariant in breast cancer as well as under various sample and process conditions, making them useful for normalizing for extraneous effects. Reference-normalized expression measurements typically range from 0 to 15, where a one unit increase generally reflects a 2-fold increase in RNA quantity.

1.4.4 Cancer-Related Genes and Reference Genes.

Table 1. Gene Expression Panel

Cancer-Related Genes	Accession Number	Reference Genes	Accession Number
BAG1	NM_004323	ACTB	NM_001101
BCL2	NM_000633	GAPD	NM_002046
CCNB1	NM_031966	GUSB	NM_000181
CD68	NM_001251	RPLP0	NM_001002
SCUBE2	NM_020974	TFRC	NM_003234
CTSL2	NM_001333		
ESR1	NM_000125		
GRB7	NM_005310		
GSTM1	NM_000561		
ERBB2	NM_004448		
MKI67	NM_002417		
MYBL2	NM_002466		
PGR	NM_000926		
STK6	NM_003600		
MMP11	NM_005940		
BIRC5	NM_001168		

1.4.5 Recurrence Score

5 The Recurrence Score (RS) on a scale from 0 to 100 is derived from the reference-normalized expression measurements as follows:

$$\begin{aligned}
 \text{RSu} &= 0.47 \times \text{GRB7 Group Threshold Score} \\
 \text{(recurrence score unscaled)} &= - 0.34 \times \text{ESR1 Group Score} \\
 &+ 1.04 \times \text{Proliferation Group Threshold Score} \\
 &+ 0.10 \times \text{Invasion Group Score} \\
 &+ 0.05 \times \text{CD68} \\
 &- 0.08 \times \text{GSTM1} \\
 &- 0.07 \times \text{BAG1}
 \end{aligned}$$

where:

$$\text{GRB7 Group Score} = 0.9 \times \text{GRB7} + 0.1 \times \text{ERBB2}$$

$$\text{GRB7 Group Threshold Score} = \begin{cases} 8 & \text{If GRB7 Group Score} < 8 \\ \text{GRB7 Group Score} & \text{Otherwise} \end{cases}$$

$$\text{ESR1 Group Score} = (\text{ESR1} + \text{PGR} + \text{BCL2} + \text{SCUBE2})/4$$

$$\text{Proliferation Group Score} = (\text{BIRC5} + \text{MKI67} + \text{MYBL2} + \text{CCNB1} + \text{STK6})/5$$

$$\text{Proliferation Group Threshold Score} = \begin{cases} 6.5 & \text{If Prolif. Group Score} < 6.5 \\ \text{Proliferation Group Score} & \text{Otherwise} \end{cases}$$

$$\text{Invasion Group Score} = (\text{CTSL2} + \text{MMP11})/2$$

The RS_u (Recurrence Score unscaled) is then rescaled to be between 0 and 100:

$$RS = \begin{cases} 0 & \text{if } 20 \times (RS_u - 6.7) < 0 \\ 20 \times (RS_u - 6.7) & \text{if } 0 \leq 20 \times (RS_u - 6.7) \leq 100 \\ 100 & \text{if } 20 \times (RS_u - 6.7) > 100 \end{cases}$$

Based on a Cox proportional hazards model, the correlation between cancer
 5 recurrence and Recurrence Score was evaluated as a continuous variable. The evaluation additionally included the Proliferation Group, GRB7 Group, ESR1 Group, Invasion Group, and each of the sixteen cancer-related genes as continuous variables.

For the purpose of identifying prognostic genes, the primary objective was to explore
 the relation between gene expression and distant recurrence-free survival (DRFS) and overall
 10 survival (OS) in the untreated patient arm. DRFS was based on the time (in years) from surgery to first distant recurrence where contralateral disease, other second primary cancers, and deaths prior to distant recurrence were considered censoring events and ipsilateral breast recurrence, local chest wall recurrence and regional recurrences were ignored. Main effect
 Cox proportional hazard models (D. R. Cox (1972) Regression Models and Life-Tables (with
 15 discussion). *J Royal Statistical Soc. B*, 34:187-220) were utilized to compare the additional contribution of gene expression beyond standard clinical prognostics variables, including age, clinical tumor size, and tumor grade. A test for comparing the reduced model, excluding the gene expression variable, versus the competing full model including the gene variable of
 interest, called the likelihood ratio test (Ronald Fisher (1922) "On the Mathematical
 20 Foundations of Theoretical Statistics", *Phil. Trans. Royal Soc.*, series A, 222:326, 1922; Leonard Savage (1962), *The Foundations of Statistical Inference* (1962)) was utilized to identify statistically significant prognostic genes.

For the purpose of identifying treatment predictive genes in breast cancer, our primary
 objective was to explore the relation between gene expression and DRFS and OS in treated
 25 patients. For such analyses, data from both treated and untreated patient arms were utilized in order to discriminate treatment predictive genes from purely prognostic genes. For identifying treatment predictive genes for tamoxifen (TAM) response, both placebo and TAM-treated patients were included from NSABP Study B-14. In both studies, Cox
 proportional hazards models were utilized to examine the interaction between the treatment
 30 effect and gene expression. An interaction between treatment and gene expression exists if

the treatment effect depends on the gene expression level; that is, if gene expression is a treatment predictive factor. Again, the likelihood ratio test was used to identify statistically significant predictive treatment genes by comparing the reduced model excluding the gene expression by treatment interaction versus the competing full model including the gene expression by treatment interaction.

Results

Table 2 reports the hazard ratio (H.R.) for recurrence versus variation in expression of genes and gene groups (axes), from analysis of the untreated NSABP B-14 trial patients (placebo arm).

Table 2. Univariate Analysis B14 Placebo Arm

Variable	Estimate	P-value (LR)	H.R.	HRLowerCL	HRUpperCL
ProlGroup	0.424845	0.0005	1.529353	1.204024	1.942588
CCNB1_2	0.437596	0.0007	1.548979	1.204938	1.991252
BIRC5_2	0.288646	0.0009	1.334619	1.12784	1.57931
MYBL2_1	0.247787	0.0026	1.281188	1.090383	1.505381
PGR_6	-0.11734	0.0075	0.889281	0.8118	0.96829
STK6_2	0.349493	0.0079	1.418348	1.098354	1.831569
MKI67_2	0.238862	0.02	1.269803	1.03654	1.55556
GSTM1_1	-0.15	0.032	0.854961	0.742047	0.985056
GAPD_1	0.305394	0.044	1.35716	1.00396	1.834618
InvasionGroup	0.253247	0.053	1.288201	0.994238	1.669079
RPLP0_2	-0.51811	0.056	0.595643	0.349143	1.016176
ProlThres	0.52227	0.061	1.685851	1.008598	2.817866
MMP11_3	0.13374	0.083	1.143096	0.982561	1.329859

As shown in Table 2, thirteen variables (genes plus gene groups) correlated with recurrence H.R. with a $p < 0.1$. As described above, because these correlations pertain to untreated patients, these variables are therefore statistically significant prognostic factors. The prognostic variables are: Proliferation Group; CCNB1; BIRC5; MYBL2; PGR; STK6; MKI67; GSTM1; GAPD; Invasion Group; RPLP0; Proliferation Threshold; MMP11. The Proliferation Group, Invasion Group, and Proliferation Threshold are defined components of the Recurrence Score algorithm.

Increased expression of ten of these prognostic factors correlates with increased H.R.: Proliferation Group; CCNB1; BIRC5; MYBL2; STK6; MKI67; GAPD; Invasion Group; Proliferation threshold; MMP11. Seven of the ten markers for poor prognosis are genes or gene sets that mark proliferating cells. Among these, the Proliferation Group (as defined in the Recurrence Score algorithm) is the top variable with respect to P-value. Increased expression three of the thirteen prognostic factors correlates with decreased H.R.: PGR, GSTM1, and RPLP0.

The results shown in Table 2 are significant not only statistically but also in their H.R. magnitudes. It should be noted that for any marker a two fold change in H.R. from 1.0 changes the H.R. by 50%, and that each HR value describes the impact of a two fold change in expression of the marker variable relative to the average expression in the patient population. Thus, for example, Table 2 shows that for every two fold increase in expression

of Proliferation Group or CCNB1 H.R. increases by approximately 50% (95% confidence limits span ~20% to ~90%).

Table 3 reports interaction analysis using results from both the placebo and TAM treatment arms of NSABP B14, carried out to identify variables that predict sensitivity or resistance to TAM.

Table 3. B14 Placebo and TAM Combined Analysis of Interaction (Int)

Variable	Estimate	P-value	H.R.	95% CI for H.R.	
IntESR1_1	-0.29602	0.000466	0.743775	0.630144	0.877896
IntSCUBE2_2	-0.21592	0.004183	0.805796	0.695108	0.934111
IntESR1 Group	-0.27804	0.006669	0.757266	0.619458	0.925732
IntTFRC_3	0.664542	0.027624	1.943601	1.075949	3.510933
IntBCL2_2	-0.26181	0.071688	0.769655	0.578846	1.023362

As shown, five variables (genes plus gene groups) correlate with response to TAM with $p < 0.1$. These are ESR1, SCUBE2, ESR1 Group, TFRC, and BCL2. The most significant of these markers is ESR1. For every two fold increase in expression of ESR1, the H.R. for recurrence of TAM-treated patients decreases by about 25% (95% confidence limits span ~12% to ~37%). These data can be used as a continuous quantitative indication of a patient's likelihood of response to TAM. This is shown graphically in Figure 1. It should be emphasized that all of the NSABP B-14 patients were classified as ER-positive, based on a clinically used assay and therefore candidates for TAM prescription. The present data demonstrate that within this "ER positive" population, patients experience variable levels of benefit on a predictable basis as a function of their ESR1 score.

This relationship between ESR1 expression and likelihood of therapeutic benefit from TAM can also be represented as a function of high, intermediate, and low ESR1 expression categories. ESR1 expression can be either represented in quartiles, tertiles or other divisions. For example, the data obtained in our study of the NSABP B14 patient population provide the following reference-normalized C_T cutpoints for division of ESR1 expression by tertiles:

Table 4. ESR1 score cutpoints that categorize ESR1 expression by tertiles,
using assay data from B-14 patients.

ER neg	0-33%ile	33-67%ile	67-100%ile
ESR1<6.5	[6.5,10.375)	[10.375,11.41)	ER≥11.41

Figure 2, shows the absolute benefit of TAM amongst these three groups of patients as a function of ESR1 Expression (Horizontal lines through bars represent 95% confidence limits.)

As shown, TAM has substantial absolute benefit in the two-thirds of patients who express ESR1 at the highest levels, but has much less impact in patients in the lowest tertile of ESR1 expression.

Similarly the relationship between ESR1 expression and TAM benefit can be represented as a function of breaking ESR1 expression into quartiles. Figure 3 shows the absolute benefit of TAM treatment as a function of quartiles of ESR1 expression (horizontal lines through bars represent 95% confidence limits). The cutpoints for ESR1 expression, are reference-normalized C_Ts derived from analysis of the NSABP B14 patient population. As shown, patients in the lowest quartile of ESR1 expression experience little benefit from TAM.

ESR1 expression data can be used in conjunction with Recurrence Score to simultaneously determine whether a patient should be prescribed TAM or TAM plus chemotherapy. Table 6, which presents data from analysis of both arms of NSABP B14, illustrates this, as well as the format for conveying these data. To illustrate how this Table might be used, patients in the low recurrence risk category as defined by RS, who express ESR1 in the upper 75th percentile are logical candidates to be treated with TAM alone. Patients who have low recurrence risk, but are in the lower 25th percentile in ESR1 expression and for whom risk of TAM side effects is a particular concern may reasonably be considered patients for whom TAM treatment is less appropriate. Patients at high risk who express ER above the 25th percentile are logical candidates for treatment with both TAM and chemotherapy.

-37-

Table 5. NSABP Patient Distributions by RS and ESR1 Expression Categories

RS Group	ER 75th-100th %ile	ER 50th-75th %ile	ER 25th-50th %ile	ER 0-25th %ile	ER neg	Row Totals
Low risk	101	89	82	41	0	313
Int risk	29	45	37	42	1	154
High Risk	24	23	32	71	28	178
All Grps	154	157	151	154	29	645

(The number within each cell is the number of assayed B-14 patients in the indicated category).

Based on the following reference-normalized ESR1 cutoff points:

Table 6

ER neg	0-25%ile	25-50%ile	50-75%ile	75-100%ile
ER<6.5	[6.5,10.065)	[10.065,10.855)	[10.855, 11.805)	ER>=11.805

The following illustrations indicate how Table 5 can be used to make decisions about treatment of a patient with anti-estrogen and/or chemotherapy.

Recommend Hormonal Therapy

5

RS Group	ER 75th-100th %ile	ER 50th-75th %ile	ER 25th-50th %ile	ER 0-25th %ile	ER neg	Row Totals
Low risk	101	89	82	41	0	313
Int risk	29	45	37	42	1	154
High Risk	24	23	32	71	28	178
All Grps	154	157	151	154	29	645

Recommend Hormonal Therapy, Consider Chemotherapy

10

15

RS Group	ER 75th-100th %ile	ER 50th-75th %ile	ER 25th-50th %ile	ER 0-25th %ile	ER neg	Row Totals
Low risk	101	89	82	41	0	313
Int risk	29	45	37	42	1	154
High Risk	24	23	32	71	28	178
All Grps	154	157	151	154	29	645

Recommend Hormonal Therapy + Chemotherapy

20

RS Group	ER 75th-100th %ile	ER 50th-75th %ile	ER 25th-50th %ile	ER 0-25th %ile	ER neg	Row Totals
Low risk	101	89	82	41	0	313
Int risk	29	45	37	42	1	154
High Risk	24	23	32	71	28	178
All Grps	154	157	151	154	29	645

All references cited throughout the disclosure are hereby expressly incorporated by reference.

25 One skilled in the art will recognize numerous methods and materials similar or equivalent to those described herein, which could be used in the practice of the present invention. Indeed, the present invention is in no way limited to the methods and materials described. While the present invention has been described with reference to what are

-39-

considered to be the specific embodiments, it is to be understood that the invention is not limited to such embodiments. To the contrary, the invention is intended to cover various modifications and equivalents included within the spirit and scope of the appended claims. For example, while the disclosure is illustrated by identifying genes and groups of genes
5 useful in predicting the beneficial response of a breast cancer patient to treatment with TAM, similar methods to determine patient response to treatment with other anti-estrogen drugs, as well as similar genes, gene sets and methods concerning other types of cancer are specifically within the scope herein.

Table 7

Reagent	Gene	Accession	Oligo	Sequence	Length
Forward	ACTB	NM_001101	S0034/B-acti.f2	CAGCAGATGTGGATCAGCAAG	21
Reverse	ACTB	NM_001101	S0036/B-acti.r2	GCATTTGCGGTGGACGAT	18
Probe	ACTB	NM_001101	S4730/B-acti.p2	AGGAGTATGACGAGTCCGGCCCC	23
Forward	BAG1	NM_004323	S1386/BAG1.f2	CGTTGTCAGCACTTGGAATACAA	23
Reverse	BAG1	NM_004323	S1387/BAG1.r2	GTTCAACCTCTTCCTGTGGACTGT	24
Probe	BAG1	NM_004323	S4731/BAG1.p2	CCCAATTAACATGACCCGGCAACCAT	26
Forward	BCL2	NM_000633	S0043/Bcl2.f2	CAGATGGACCTAGTACCCACTGAGA	25
Reverse	BCL2	NM_000633	S0045/Bcl2.r2	CCTATGATTAAAGGGCATTTTTCC	24
Probe	BCL2	NM_000633	S4732/Bcl2.p2	TTCCACGCCGAAGGACAGCGAT	22
Forward	CCNB1	NM_031966	S1720/CCNB1.f2	TTCAGGTTGTTGCAGGAGAC	20
Reverse	CCNB1	NM_031966	S1721/CCNB1.r2	CATCTTCTTGGGCACACAAT	20
Probe	CCNB1	NM_031966	S4733/CCNB1.p2	TGTCTCCATTATTGATCGGTTTCATGCA	27
Forward	CD68	NM_001251	S0067/CD68.f2	TGGTTCCCAGCCCTGTGT	18
Reverse	CD68	NM_001251	S0069/CD68.r2	CTCCTCCACCCTGGGTTGT	19
Probe	CD68	NM_001251	S4734/CD68.p2	CTCCAAGCCCAGATTTCAGATTCGAGTCA	28
Forward	SCUBE2	NM_020974	S1494/SCUBE2.f2	TGACAATCAGCACACCTGCAT	21
Reverse	SCUBE2	NM_020974	S1495/SCUBE2.r2	TGTGACTACAGCCGTGATCCTTA	23
Probe	SCUBE2	NM_020974	S4735/SCUBE2.p2	CAGGCCCTCTTCCGAGCGGT	20
Forward	CTSL2	NM_001333	S4354/CTSL2.f1	TGTCTCACTGAGCGAGCAGAA	21
Reverse	CTSL2	NM_001333	S4355/CTSL2.r1	ACCATTGCAGCCCTGATTG	19
Probe	CTSL2	NM_001333	S4356/CTSL2.p1	CTTGAGGACGCGAACAGTCCACCA	24
Forward	ESR1	NM_000125	S0115/EstR1.f1	CGTGGTGCCCTCTATGAC	19
Reverse	ESR1	NM_000125	S0117/EstR1.r1	GGCTAGTGGGCGCATGTAG	19
Probe	ESR1	NM_000125	S4737/EstR1.p1	CTGGAGATGCTGGACGCCC	19
Forward	GAPD	NM_002046	S0374/GAPD.f1	ATCCACCCATGGCAAATTC	20
Reverse	GAPD	NM_002046	S0375/GAPD.r1	GATGGGATTTCCATTGATGACA	22
Probe	GAPD	NM_002046	S4738/GAPD.p1	CCGTTCTCAGCCTTGACGGTGC	22
Forward	GRB7	NM_005310	S0130/GRB7.f2	CCATCTGCATCCATCTTGTT	20
Reverse	GRB7	NM_005310	S0132/GRB7.r2	GGCCACCAGGGTATTATCTG	20
Probe	GRB7	NM_005310	S4726/GRB7.p2	CTCCCCACCCTTGAGAAAGTGCCT	23
Forward	GSTM1	NM_000561	S2026/GSTM1.r1	GGCCCAGCTTGAATTTTCA	20
Reverse	GSTM1	NM_000561	S2027/GSTM1.f1	AAGCTATGAGGAAAAGAAGTACACGAT	27
Probe	GSTM1	NM_000561	S4739/GSTM1.p1	TCAGCCACTGGCTTCTGTCATAATCAGGAG	30
Forward	GUSB	NM_000181	S0139/GUS.f1	CCCACTCAGTAGCCAAGTCA	20
Reverse	GUSB	NM_000181	S0141/GUS.r1	CACGCAGGTGGTATCAGTCT	20
Probe	GUSB	NM_000181	S4740/GUS.p1	TCAAGTAAACGGGCTGTTTTCCAAACA	27
Forward	ERBB2	NM_004448	S0142/HER2.f3	CGGTGTGAGAAGTGCAGCAA	20
Reverse	ERBB2	NM_004448	S0144/HER2.r3	CCTCTCGCAAGTGCTCCAT	19
Probe	ERBB2	NM_004448	S4729/HER2.p3	CCAGACCATAGCACACTCGGGCAC	24
Forward	MKI67	NM_002417	S0436/MKI67.f2	CGGACTTTGGGTGCGACTT	19
Reverse	MKI67	NM_002417	S0437/MKI67.r2	TTACAACCTCTTCCACTGGGACGAT	24
Probe	MKI67	NM_002417	S4741/MKI67.p2	CCACTTGTGGAACCAACCGCTCGT	23
Forward	MYBL2	NM_002466	S3270/MYBL2.f1	GCCGAGATCGCCAAGATG	18
Reverse	MYBL2	NM_002466	S3271/MYBL2.r1	CTTTTGATGGTAGAGTTCCAGTGATTG	27
Probe	MYBL2	NM_002466	S4742/MYBL2.p1	CAGCATTGTCTGTCCTCCCTGGCA	24
Forward	PGR	NM_000926	S1336/PR.f6	GCATCAGGCTGTCATTATGG	20
Reverse	PGR	NM_000926	S1337/PR.r6	AGTAGTTGTGCTGCCCTTCC	20
Probe	PGR	NM_000926	S4743/PR.p6	TGTCCTTACCTGTGGGAGCTGTAAGGTC	28
Forward	RPLP0	NM_001002	S0256/RPLP0.f2	CCATTCTATCATCAACGGGTACAA	24
Reverse	RPLP0	NM_001002	S0258/RPLP0.r2	TCAGCAAGTGGGAAGGTGTAATC	23

-41-

Probe	RPLP0	NM_001002	S4744/RPLPO.p2	TCTCCACAGACAAGGCCAGGACTCG	25
Forward	STK6	NM_003600	S0794/STK6.f2	CATCTTCCAGGAGGACCACT	20
Reverse	STK6	NM_003600	S0795/STK6.r2	TCCGACCTTCAATCATTTCA	20
Probe	STK6	NM_003600	S4745/STK6.p2	CTCTGTGGCACCCTGGACTACCTG	24
Forward	MMP11	NM_005940	S2067/MMP11.f3	CCTGGAGGCTGCAACATACC	20
Reverse	MMP11	NM_005940	S2068/MMP11.r3	TACAATGGCTTTGGAGGATAGCA	23
Probe	MMP11	NM_005940	S4746/MMP11.p3	ATCCTCCTGAAGCCCTTTTCGCAGC	25
Forward	BIRC5	NM_001168	S0259/BIRC5.f2	TGTTTTGATTCCCGGGCTTA	20
Reverse	BIRC5	NM_001168	S0261/BIRC5.r2	CAAAGCTGTCAGCTCTAGCAAAAG	24
Probe	BIRC5	NM_001168	S4747/BIRC5.p2	TGCCTTCTTCCTCCCTCACTTCTCACCT	28
Forward	TFRC	NM_003234	S1352/TFRC.f3	GCCAACTGCTTTCATTTGTG	20
Reverse	TFRC	NM_003234	S1353/TFRC.r3	ACTCAGGCCCCATTTCTTTA	20
Probe	TFRC	NM_003234	S4748/TFRC.p3	AGGGATCTGAACCAATACAGAGCAGACA	28

Table 8

Gene	LocusLink	Sequence
ACTB	NM_001101	CAGCAGATGTGGATCAGCAAGCAGGAGTATGACGAGTCCGGCCCTCCATCGTCCACCGCAATGC
BAG1	NM_004323	CGTTGTCAGCACTTGGATACAAGATGGTTGCCGGTCATGTTAATTGGGAAAAAGAACAGTCCACAGGAAGAGTTGAAC
BCL2	NM_000633	CAGATGGACCTAGTACCCACTGAGATTTCACAGCGCGAAGGACAGCGATGGGAAATGCCCTTAATCATAGG
CCNB1	NM_031966	TTCAGGTTGTTCAGGAGACCATTGATGACTGTCTCCATTATTGATCGGTTTCATGCAGATAATTGTGTGCCCAAGAAGATG
CD68	NM_001251	TGGTTCACGCCCCTGTGTCCACCTCCAGCCAGATTTCAGATTGAGTCAATGTACACAAACCAGGGTGGAGGAG
SCUBE2	NM_020974	TGACAAATCAGCACACCTGCATTACCGCTCGGAAGAGGGCCCTGAGCTGCATGAATAAGGATCAGGGCTGTAGTCACA
CTSL2	NM_001333	TGTCTCACTGAGCGAGCAGAAATCTGGTGGACTGTTCCGCTCCTCAAGGCAATCAGGGCTGCAATGGT
ESR1	NM_000125	CGTGGTGCCTCTATGACCTGCTGCTGGAGATGCTGGACGCCACCGCTACATCGGCCCACTAGCC
GAPD	NM_002046	ATTCCACCCATGGCAAAATCCATGGCACCCGTCAAGGCTGAGAACGGGAAGCTTGTCATCAATGGAAATCCCATC
GRB7	NM_005310	CCATCTGCATCCATCTTTGGGCTCCACCCCTTGAGAAAGTGCCCTCAGATAATACCCCTGGTGGCC
GSTM1	NM_000561	AAGCTATGAGGAAAGAAAGTACACGATGGGGACGCTCCTGATTATGACAGAAAGCCAGTGGCTGAATGAAAATTCAAGCTGGGCC
GUSB	NM_000181	CCCACTCAGTAGCCCAAGTCACAATGTTTGGAAACAGCCCGTTTACTTGAGCAAGACTGATACCACCTGGGTG
ERBB2	NM_004446	CGGTGTGAGAAAGTGCAGCAAGCCCTGTGCGGAGTGTGCTATGGTCTGGGCATGGAGCACTTGGCGAGAGG
MKI67	NM_002417	CGGACTTTGGGTGCGACTTGACGAGCGGTGTTCGACAAAGTGGCCCTTGGCGGCGGATCGTCCAGTGGGAAGTTGTAA
MYBL2	NM_002466	GCCGAGATCGCCAAAGATGTTGCCAGGGAGGACAGACAAATGCTGTGAAGAATCACTGGAACCTCTAGCATCAAAAG
PGR	NM_000926	GCATCAGGCTGTCAATTATGGTGTCTTACCTGTGGGAGCTGTAAAGGTCTTCTTTAAGAGGGCAATGGAGGCGACACAATACT
RPLP0	NM_001002	CCATTCTATCATCAACGGGTACAAACGAGTCTTGGCCCTTGTCTGTGGAGACGGATTACACCTTCCCACTTGTCTGA
STK6	NM_003600	CATCTTCCAGGAGAACCACTCTCTGTGGCAACCCCTGGACTACCTGCCCCCTGAAATGATTGAAGTGCGGA
MMP11	NM_005940	CCTGGAGGCTGCAACATACCTCAATCCTGTCCAGGCCGGATCCTCTGAAAGCCCTTTTCGCAGCACTGCTATCCTCCAAAGCCATTGTA
BIRC5	NM_001168	TGTTTTCATTCCCGGCTTACGAGGTGAGAAGTGAGGGAGGAAGGCAAGTGTCCCTTTTGTAGAGCTGACAGCTTIG
TFR3	NM_003234	GCCAACTGCTTTCATTGTGAGGGATCTGAACCAATAACAGAGCAGACATAAAGGAATGGGCCCTGAGT

WHAT IS CLAIMED:

1. A method for the prognosis of disease outcome in a breast cancer patient, comprising

(a) quantitatively determining, in a biological sample comprising cancer cells obtained from said patient, the value of one or more of the following variables:

(i) Proliferation Group Score;

(ii) Invasion Group Score;

(iii) Proliferation Group Threshold Score; and

(iv) the expression levels of the RNA transcripts of one or more of the following individual genes CCNB1, BIRC5, MYBL2, PGR, STK6, MKI67, GSTM1, GAPD, RPLPO, and MMP11, or their expression products; wherein

(b1) for every unit of an increase in the value of one or more of (i) - (iii) and/or of the RNA transcript(s) of one or more of the individual genes CCNB1, BIRC5, MYBL2, STK6, MKI67, GAPD, and MMP11, or the corresponding expression product(s), said patient is identified to have a proportionately increased risk of poor disease outcome; and

(b2) for every unit of increased expression level(s) of the RNA transcript(s) of one or more of the individual genes PGR, GSTM1, and RPLPO, or the corresponding expression product(s), said patient is identified to have a proportionately decreased risk of poor disease outcome,

wherein

Proliferation Group Score = $(BIRC5 + MKI67 + MYBL2 + CCNB1 + STK6)/5$;

Invasion Group Score = $(CTSL2 + MMP11)/2$;

Proliferation Group Threshold Score equals 6.5, if the Proliferation Group Score is less than 6.5; and is identical with the Proliferation Group Score, if the Proliferation Group Score is 6.5 or more, wherein

the gene symbols in the equations represent the expression levels of the RNA transcripts of the respective genes, or their expression products; and

wherein every individual gene or gene present in any of said variables can be substituted by another gene that coexpresses in the applicable tumor type with said gene in said cancer with a Pearson's coefficient of ≥ 0.5 .

2. The method of claim 1, wherein said patient is a human patient.

3. The method of claim 2, wherein said disease outcome is expressed as overall patient survival.
4. The method of claim 2, wherein said disease outcome is expressed as recurrence-free survival.
5. The method of claim 2, wherein said disease outcome is expressed as distant recurrence-free survival.
6. The method of claim 2 wherein said prognosis assumes that the patient receives no further treatment after surgical resection of said breast cancer.
7. The method of claim 2 wherein said expression levels are normalized relative to the expression levels of one or more reference genes, or their expression products.
8. The method of claim 7 wherein said reference genes are selected from the group consisting of ACTB, GAPD, GUS, RPLPO, and TFRC.
9. The method of claim 7 wherein said expression levels are normalized relative to the mean of the expression levels of ACTB, GAPD, GUS, RPLPO, and TFRC.
10. The method of claim 2 wherein the quantitative value of said disease outcome is directly proportionate with the value of the variable or variables determined over a continuum.
11. The method of claim 2 comprising the determination of the Proliferation Score.
12. The method of claim 11 further comprising the determination of the Proliferation Group Threshold Score.
13. The method of claim 12 further comprising determination of the Invasion Group Score.

14. The method of claim 2 comprising determination of the expression levels of at least two of the individual genes listed in (iv), or their expression products.

15. The method of claim 2 comprising determination of the expression levels of at least three of the individual genes listed in (iv), or their expression products.

16. The method of claim 2 comprising determination of the expression levels of at least four of the individual genes listed in (iv), or their expression products.

17. The method of claim 2 comprising determination of the expression levels of at least five of the individual genes listed in (iv), or their expression products.

18. The method of claim 2 comprising determination of the expression levels of at least six of the individual genes listed in (iv), or their expression products.

19. The method of claim 2 comprising determination of the expression levels of at least seven of the individual genes listed in (iv), or their expression products.

20. The method of claim 2 comprising determination of the expression levels of at least eight of the individual genes listed in (iv), or their expression products.

21. The method of claim 2 comprising determination of at least nine of the expression levels of the individual genes listed in (iv), or their expression products.

22. The method of claim 2 comprising determination of the expression levels of all individual genes listed in (iv), or their expression products.

23. The method of claim 2 wherein said breast cancer is lymph node negative.

24. The method of claim 23 wherein said breast cancer is ESR1 positive.

25. The method of claim 2 wherein said determination is performed more than one time.

26. The method of claim 25 wherein said determination is performed prior to and following chemotherapy, hormonal therapy and/or radiation therapy.

27. The method of claim 2 wherein said patient is determined to have an increased risk of poor disease outcome.

28. The method of claim 27 wherein following said determination said patient is treated with chemotherapy, hormonal therapy and/or radiation therapy.

29. The method of claim 28 wherein said chemotherapy is neoadjuvant chemotherapy.

30. The method of claim 29 wherein said chemotherapy comprises the administration of a taxane derivative.

31. The method of claim 30 wherein said taxane derivative is docetaxel or paclitaxel.

32. The method of claim 31 wherein said taxane derivative is docetaxel.

33. The method of claim 29 wherein said chemotherapy comprises the administration of an anthracycline derivative.

34. The method of claim 33 wherein said anthracycline derivative is doxorubicin.

35. The method of claim 29 wherein said chemotherapy comprises the administration of a topoisomerase inhibitor.

36. The method of claim 35 wherein said topoisomerase inhibitor is selected from the group consisting of camptothecin, topotecan, irinotecan, 20-S-camptothecin, 9-nitro-camptothecin, 9-amino-camptothecin, and GI147211.

37. The method of claim 28 wherein said hormonal therapy comprises the administration of TAM.

38. The method of claim 28 wherein said hormonal therapy comprises the administration of an anti-estrogen drug selected from the group consisting of toremifene, anastrozole, and megesterol acetate.

39. The method of claim 1 wherein said biological sample is a tissue sample comprising cancer cells.

40. The method of claim 39 wherein said tissue is fixed, paraffin-embedded, or fresh, or frozen.

41. The method of claim 40 where the tissue is from fine needle, core, or other types of biopsy.

42. The method of claim 40 wherein the tissue sample is obtained by fine needle aspiration, bronchial lavage, or transbronchial biopsy.

43. The method of claim 1 wherein said expression levels are determined by quantitative RT-PCR.

44. The method of claim 1 wherein the expression level of said expression product or products is determined by immunohistochemistry.

45. The method of claim 1 wherein the expression level of said expression product or products is determined by proteomics techniques.

-48-

46. The method of claim 2 further comprising the step of creating a report summarizing said prognosis.

47. A method for quantitative determination of the likelihood of a beneficial response of an ESR1 positive breast cancer patient to treatment with an anti-estrogen drug, comprising quantitatively determining, in a biological sample comprising cancer cells obtained from said patient, one or more of the following variables:

- (i) ESR1 Group Score; and
- (ii) the expression levels of the RNA transcripts of one or more of the following individual genes ESR1, SCUBE2, TFRC, and BCL2, or their expression products;

wherein

for every unit of increased numerical value of the ESR1 group score, ESR1, SCUBE2, or BCL2 variables said patient is identified to have a proportionately increased likelihood of a beneficial response to treatment with an anti-estrogen drug, and for every unit of increased numerical value of the TFRC variable said patient is identified to have a proportionately decreased likelihood of a beneficial response to treatment with an anti-estrogen drug.

48. The method of claim 47 wherein the ESR1 group score, or the expression level of the ESR1 gene or its expression product is determined.

49. The method of claim 48 wherein the expression level of the ESR1 gene or its expression product is determined.

50. The method of claim 47 wherein the anti-estrogen drug is selected from the group consisting of tamoxifen, toremifene, anastrozole, and megesterol acetate.

51. The method of claim 50 wherein the anti-estrogen drug is tamoxifen.

52. The method of claim 47 further comprising the step of preparing a report for said patient, including a treatment recommendation.

53. The method of claim 52 wherein treatment with anti-estrogen alone, chemotherapy alone, or chemotherapy plus anti-estrogen is recommended by consulting a

table or graph along one axis of which ESR1 expression level is displayed, either as a continuous variable or in ESR1 expression ranges, and along the other axis of which likelihood of cancer recurrence is displayed, either as a continuous variable or in risk ranges.

54. The method of claim 53 wherein said expression ranges are quartiles within a population of breast cancer patient ESR1 values.

55. The method of claim 53 wherein said risk ranges are low, intermediate and high risk ranges.

56. The method of claim 53 wherein said patients in the high ESR1 and low risk categories are administered anti-estrogen alone.

57. The method of claim 53 wherein said patients in the low ESR1 and high risk category are administered chemotherapy alone.

58. The method of claim 53 wherein said patients in the intermediate ESR1 and high risk category are administered both anti-estrogen and chemotherapy.

59. The method of claim 53 wherein a patient is known to be in a low risk group on the basis of a diagnostics test based on the expression levels of more than five genes.

60. The method of claim 59 where the test is measured the Recurrence Score.

61. The method of claim 53 wherein said variable is the expression level of the RNA transcript of ESR1, or its expression product.

62. The method of claim 47 further comprising the step of determining the Recurrence Score for said patient.

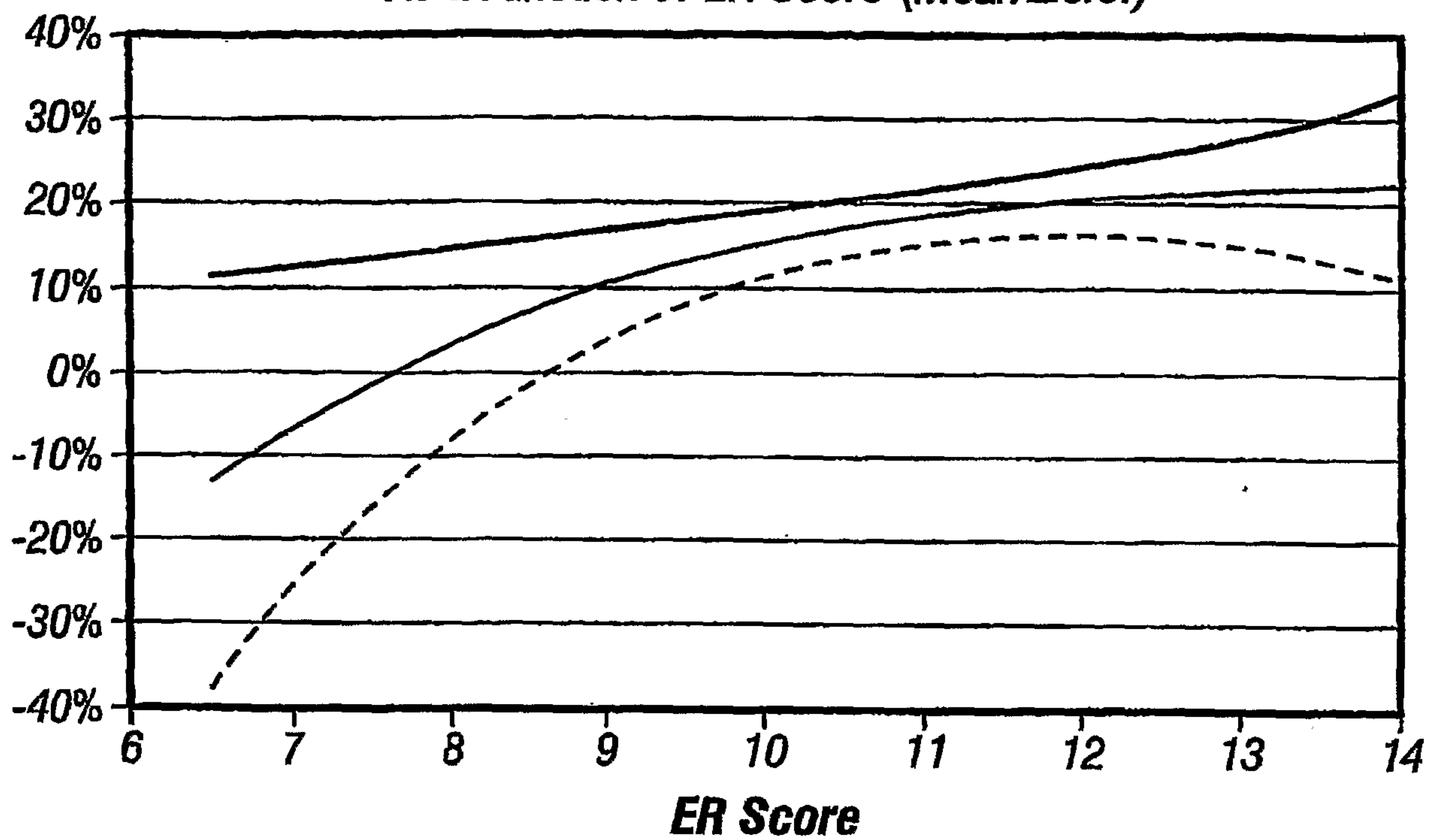
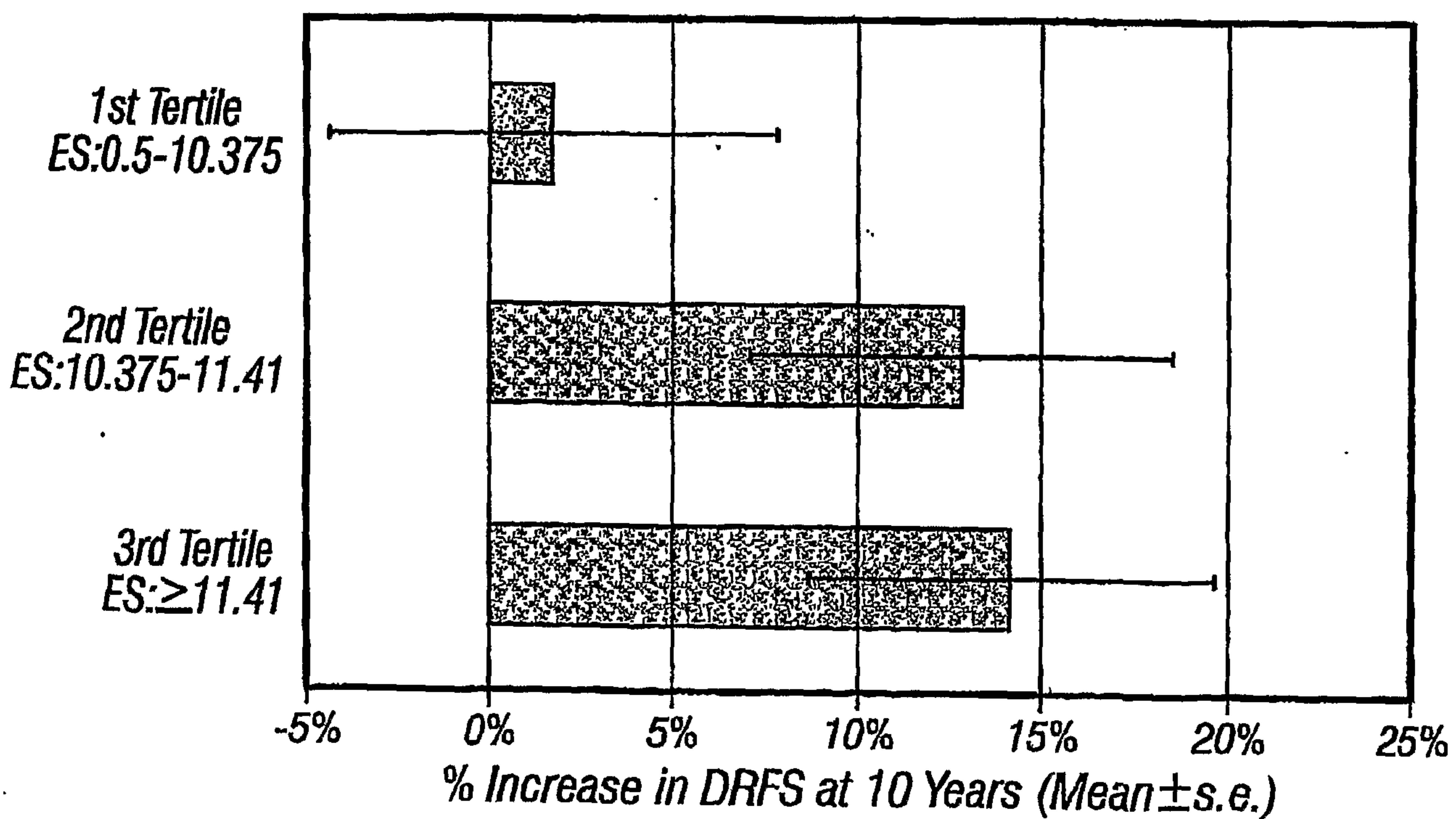
63. The method of claim 62 wherein the anti-estrogen drug is selected from the group consisting of tamoxifen, toremifene, anastrozole, and megesterol acetate.

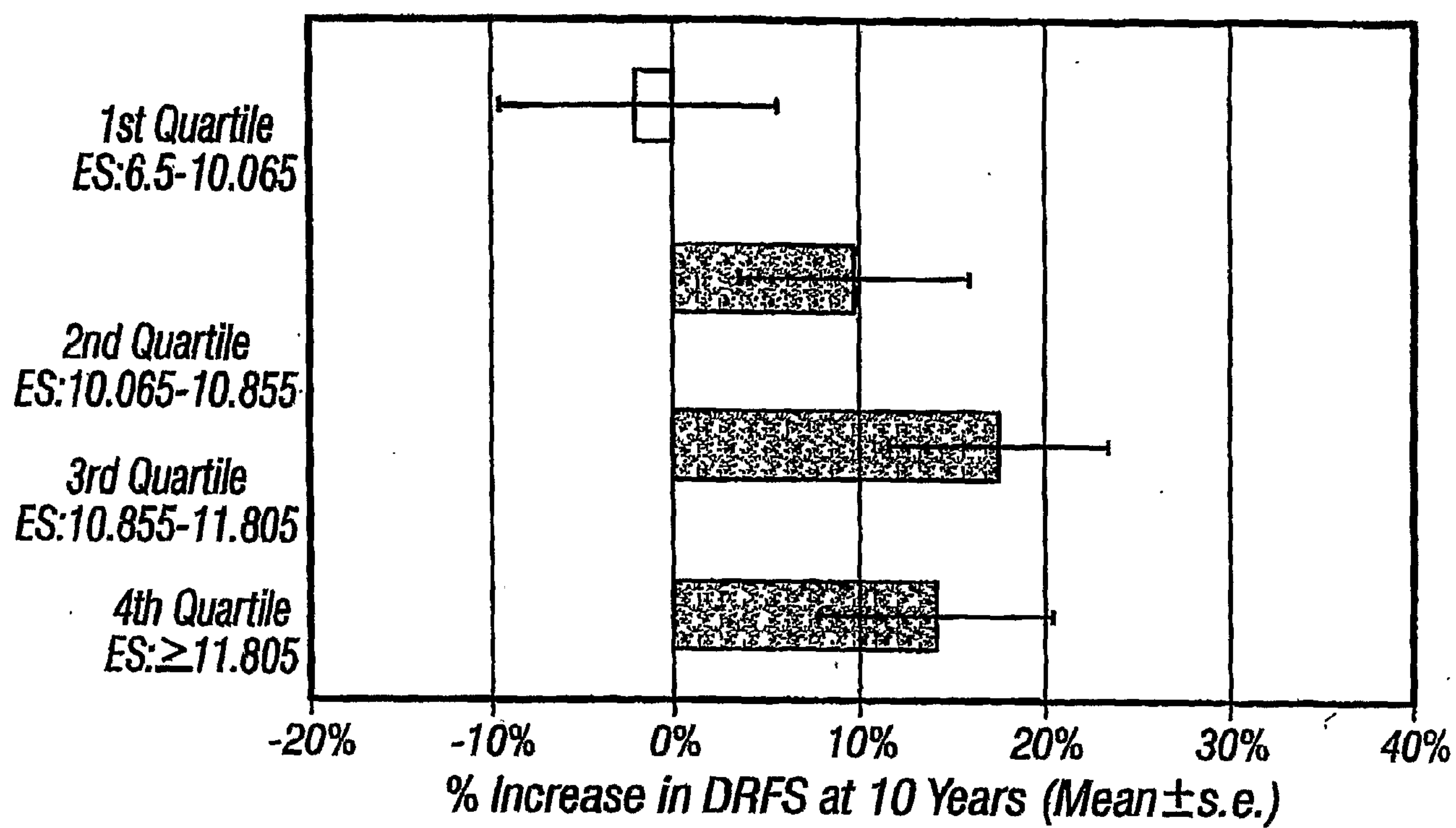
64. The method of claim 63 wherein the anti-estrogen drug is tamoxifen.
65. The method of claim 62 further comprising the step of preparing a report for said patient, including a treatment recommendation. Based on claim 53.
66. A kit comprising a set of gene specific probes and/or primers for quantifying the expression of one or more of the genes listed in Table 1 by quantitative RT-PCR.
67. The kit of claim 66 wherein the gene specific probes are selected from the group consisting of the probes listed in Table 7.
68. The kit of claim 66 wherein the gene specific primers are selected from the group consisting of the forward and reverse primers listed in Table 7.
69. The kit of claim 66 wherein the amplicon used in said quantitative RT-PCR is selected from the group consisting of the amplicons listed in Table 8.
70. The kit of claim 66 further comprising one or more reagents for expression of RNA from tumor samples.
71. The kit of claim 66 comprising one or more containers.
72. The kit of claim 66 comprising one or more algorithms that yield prognostic or predictive information.
73. The kit of claim 71 wherein one or more of said containers comprise pre-fabricated microarrays, a buffers, nucleotide triphosphates, reverse transcriptase, DNA polymerase, RNA polymerase, probes, or primers.
74. The kit of claim 66 comprising a label or package insert with instructions for use of its components.

75. The kit of claim 74 wherein said instructions comprise directions for use in the prediction or prognosis of breast cancer.

1/2

***Absolute Benefit of Tamoxifen
As a Function of ER Score (Mean $\pm$ s.e.)***

**FIG. 1****FIG. 2**

2/2**FIG. 3**

*Absolute Benefit of Tamoxifen
As a Function of ER Score (Mean $\pm$ s.e.)*

